

# Implementing a Voice Effects ACASI Experiment: Considerations and Preliminary Effects

Karen Jaques

UW Survey Center  
University of Wisconsin Madison  
kjaques@ssc.wisc.edu

Paper presented at the annual meetings of the  
International Field Directors and Technology Conference  
Pasadena, CA  
18-21 May 2014

© 2014. Materials may not be reproduced without permission of the author.

# Overview

---

- We implemented a complex experiment into the ACASI section of a CAPI study of 17-year old foster youth, residing in California.
- This presentation will focus on:
  - Set up and implementation the experiment
  - Using a text-to-speech voice
  - Preliminary look at the data

# ACASI (Audio Computer-Assisted Self-Interview)

---

- For an ACASI section, the interviewer hands over the laptop to the respondent who listens to the pre-recorded questions using headphones.
- The respondent will listen and answer the questions independently. Once the ACASI section is over, the interviewer takes back the laptop and administers the rest of the interview to the respondent.

# ACASI (Audio Computer-Assisted Self-Interview)

---

- ACASI is the preferred method for administering sensitive questions in face-to-face interviews because:
  - it often yields higher reports of sensitive behaviors compared to CAPI or paper-and-pencil questionnaires (Tourangeau and Smith 1996; Turner, Ku, et al. 1998).
  - the inclusion of audio can overcome problems respondents might have because of literacy (Turner, Forsyth et al. 1998).

# ACASI (Audio Computer-Assisted Self-Interview)

---

- However, a small body of research raises doubts about:
  - (1) whether respondents actually use the audio; and
  - (2) whether the inclusion of audio has any effect on survey responses (Couper, Tourangeau, and Marvin 2009).
- Given the potential benefits of including audio, we believe the impetus should be on researchers to determine methods for decreasing the barriers to listening to the audio.

# The CalYOUTH Study

## (California Youth Transitions to Adulthood Study)

---

- We interviewed 727 17-year old foster youth in California between April 2013 – October 2013.
- The average interview was about 100 minutes.
- Interview included questions about the youth's experiences and opinions about foster care and juvenile courts, education, employment, parenting, relationships and plans for the future.

# The CalYOUTH Study (California Youth Transitions to Adulthood Study)

---

- The ACASI portion included questions about:
  - Sex and pregnancy
  - Delinquency
  - Maltreatment prior to foster care
  - Sexual abuse
  - Mental health
  - Drug and alcohol use
- This unique sample allowed us to examine the impact of audio on reports about sensitive behaviors in a population at-risk for low literacy and for engaging in the behaviors.

# CaYOUTH ACASI Experiment Goals

---

## **Goal 1: Voluntary vs Involuntary Audio**

Experimentally test the impact of making use of the audio component voluntary versus involuntary and evaluating the effect on levels of reporting about sensitive behaviors.

## **Goal 2: Choice of Voice**

In order to increase the likelihood that respondents will refrain from turning the audio off we tested the impact of offering respondents a choice of voices.

- We expected that giving respondents a choice of voice to listen to would decrease the negative experience of listening to the audio and emphasize the importance of the audio component in the ACASI administration.



# The CalYOUTH ACASI Experiment

---

- Our original experiment implemented a 2x2 factorial design, where respondents were randomly assigned to one of four groups.

---

**Figure 1. Overview of experimental design**

---

<b>Factor 2: Voice Choice</b>	<b>Factor 1: Involuntary Audio</b>	
	Voluntary audio	Involuntary audio
Random assignment to voice	Group 1 (N=346)	Group 3 (N=47)
Option to select 1 of 3 possible voices	Group 2 (N=302)	Group 4 (N=26)

- After about a month in the field, we had to drop the involuntary audio component due to the client's concern that it was extending the overall length of the interview.

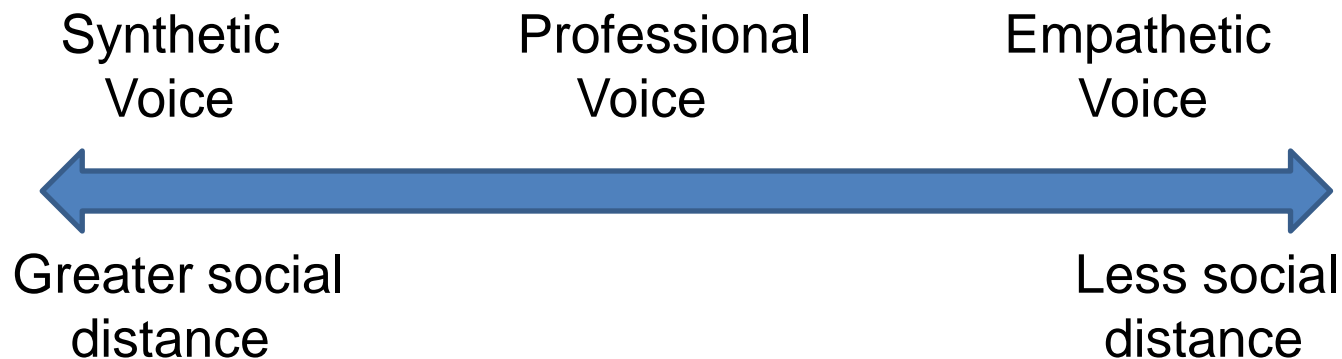
# Selecting the Voices

© 2014. Materials may not be reproduced without permission of the author

# Selecting Voices

---

- The voices were selected to represent different personas: Empathetic, Professional, Synthetic (Evans and Kortum 2009, 2010).
- These voice options for respondents were distributed along a continuum representing low to high levels of social distancing (Johnson and Parsons 1994).



# Selecting Voices

---

- We used all female voices. This decision was based on our analysis of Wave 1 of the Midwest Young Adult Study that demonstrated improved reporting for male respondents with a female voice (Dykema, DiLoreto, Price, White and Schaeffer 2012).
- Both human voices were from women that sounded in their twenties.
- We selected a text-to-speech voice that sounded most similar in age to our human voices.

# Voice Ratings

---

- We devised a formal independent voice ratings web survey, where internal staff between the ages of 18 and 25 were asked to evaluate each voice on basic demographics and also empathetic and professional characteristics (How friendly, how trustworthy, how comfortable, etc).
- We recorded samples of 6 voices (5 human; 1 synthetic) and selected the 2 human voices that most embodied the empathetic and professional characteristics based on these ratings.
- We also coached each human voice to try to enhance either the empathetic or professional characteristic of their voice prior to recording the samples.

# Examples of Voices

---

- The following audio clips were the exact files that respondents listened to before choosing a voice.

Voice 1 – Empathetic



Voice 2 – Professional



Voice 3 – Synthetic (TTS)



# Text-to-speech (TTS) Software

© 2014. Materials may not be reproduced without permission of the author

# TTS software

---

- We researched the available options (as of 2012) and chose **TextSpeech Pro 3.6.0**.
- Reasons:
  - Ability to modify speech properties (voice, speed, volume, pitch) and speech entities (pauses) on the spot.
  - Utilizes *AT&T Natural Voices* which uses a more sophisticated algorithm to make synthetic voice sound more natural or human-like.
  - Included a voice that most closely matched the demographic characteristics of our human voices.
  - Affordable software (about \$100)



## TTS Software – Drawbacks/Issues

---

- Software's sophisticated algorithm uses different phoneme sounds depending on the context. For example, it attempts to use a hard "t" or soft "t" sound depending on the placement of that sound in a word. The idea is to produce a more natural (human) sounding voice.
  - Example: **Turtle** vs. **Little**
- Unfortunately, this feature did not always work as intended. Sometimes the voice would add seemingly random emphasis to a syllable or word depending on its placement within a sentence.
- This issue required extensive audio editing before the file could be used in the study.

# TTS Software – Drawbacks/Issues

---

- **Listening Example 1**

- Sound clip A – The word “secretly” recorded within a sentence



- Sound clip B - The word “secretly” recorded alone



- Sound clip C – Edited version of the same sentence



# TTS Software – Drawbacks/Issues

---

- **Listening Example 2**

“How often were you loud, rowdy, or unruly in a public place?”

- Sound clip A – Recorded as a full sentence



- Sound clip B - The word “rowdy” recorded alone



- Sound clip C – “Rowdy” in a different sentence “How often were you rowdy?”



# Training Interviewers

© 2014. Materials may not be reproduced without permission of the author

# Implementing the Experiment in the Field

---

- Our CAPI interviewers were trained on the experimental design prior to fielding.
- Interviewers were not aware of which experiment group a respondent was assigned to until they reached the ACASI section. In this way, they needed to be ready for each option.

# Implementing the Experiment in the Field

---

- In some cases interviewers needed to walk a respondent through the voice selection screens and answer any questions the respondent might have, while not giving away to the respondent that this was an experiment.
- It was important that interviewers understood the experiment and had buy-in so that they would properly follow the protocols.

# Handling Complaints in ACASI

---

- If a respondent complained about the voice, the interviewer could intervene and change the voice for the respondent.
  - If the respondent was still unhappy, the interviewer would offer to administer the questions to the respondent.
  - Interviewers would need to take back the laptop and navigate hidden screens to comply with either scenario.

# The Instrument

© 2014. Materials may not be reproduced without permission of the author



# Programming the Instrument

---

Programming this experiment presented several challenges:

1. The need to convey the voice selection treatment to respondents so that they could easily navigate through.
2. The need to collect paradata to analyze the results of the experiment.
  - We collected information about whether the respondent turned on or off the question audio and/or the response option audio,
  - the time spent in the question, and
  - whether the voice was changed for each item. This resulted in hundreds of paradata variables.

# Voice Choice Screens for Respondent

---

QA Client - Q:\projects\p9927\prac\p9927\e-inst  
Caseid: 90038  
Item: ACASI\_V1  
Press ENTER to move to the next field

Press "enter" to listen to Voice 1.

QA Client - Q:\projects\p9927\prac\p9927\e-inst  
Caseid: 90038  
Item: ACASI\_V2

Press "enter" to listen to Voice 1.

Voice 1 is finished playing.

Press "**enter**" to listen to **Voice 2** or press "**x**" to listen to **Voice 1** again.

# Voice Choice Screens for Respondent

```
QA Client - C:\uwsprojects\p9927\e-inst
Caseid: 90004
Item: ACASI_1_VOICE@a

Press "enter" to listen to Voice 1.

Voice 1 is finished playing.
Press "enter" to listen to Voice 2 or press "x" to listen to Voice 1 again.

Voice 2 is finished playing.
Press "enter" to listen to Voice 3 or press "x" to listen to Voice 2 again.

Voice 3 is finished playing.
Press "enter" to continue or press "x" to listen to Voice 3 again.

Please select the voice you liked listening to the most.

Type 1 for the first voice, 2 for the second voice, or 3 for the third voice.

<1> SELECT VOICE 1
<2> SELECT VOICE 2
<3> SELECT VOICE 3


```

# Involuntary vs Voluntary Audio Options

QA Client - C:\uwsprojects\p9927\e-inst

Caseid: 90004  
Item: PRG3\_N\_W1

Have you ever given birth to any children?

Enter 1 for YES or 2 for NO.

<1> YES  
<2> NO  
<d> DON'T KNOW      <r> REFUSED

<q> REPLAY QUESTION      <h> REPLAY RESPONSES  
<y> TURN QUESTION AUDIO OFF - (SOUND IS NOW **ON**)  
<s> TURN RESPONSE AUDIO OFF - (SOUND IS NOW **ON**)

# Preliminary Findings

© 2014. Materials may not be reproduced without permission of the author

# Which voice did respondents choose?

---

Table 1. Voice respondent's chose to listen to

<b>Voice Choice</b>	<b>N (%)</b>
Empathetic	156 (47.4%)
Professional	162 (49.2%)
Synthetic	11 (3.3%)
TOTAL	329 (100%)

# Which voice did respondents choose?

---

Table 2. Voice respondent's chose to listen to by gender

<b>Voice Choice</b>	<b>MALE</b>	<b>FEMALE</b>	<b>TOTAL</b>
Empathetic	56 (43.1%)	100 (50.3%)	156 (47.4%)
Professional	68 (52.3%)	94 (47.2%)	162 (49.2%)
Synthetic	6 (4.6%)	5 (2.5%)	11 (3.3%)
<b>TOTAL</b>	<b>130 (100%)</b>	<b>199 (100%)</b>	<b>329 (100%)</b>

# When did respondent's turn off audio?

---

Table 3. Turned off audio by quartile

<b>Quartile</b>	<b>Groups 1 and 2 (Voluntary audio)</b>
<b>1</b>	282 (45.5%)
<b>2</b>	43 (6.6%)
<b>3</b>	11 (1.7%)
<b>4</b>	15 (2.3%)
<b>Never turned off</b>	297 (45.8%)
<b>TOTAL</b>	648 (100%)



# Turning off audio and option to select voice

---

Table 4. Turned audio off by treatment group

	<b>Group 1: Option to select voice</b>	<b>Group 2: Randomly assigned voice</b>	<b>TOTAL</b>
<b>Ever turned audio off</b>	193 (55.8%)	158 (52.3%)	351 (54.2%)
<b><u>Did not turn</u> audio off</b>	153 (44.2%)	144 (47.7%)	297 (45.8%)
<b>TOTAL</b>	346 (100%)	302 (100%)	648 (100%)

# Turning off audio and Voice Choice

---

Table 5. Turned audio off by voice choice

	<b>Empathetic Voice</b>	<b>Professional Voice</b>	<b>Synthetic Voice</b>	<b>TOTAL</b>
<b>Ever turned audio off</b>	133 (55.4%)	140 (52.0%)	78 (56.1%)	351 (54.2%)
<b><u>Did not turn</u> audio off</b>	107 (44.6%)	129 (48.0%)	61 (43.9%)	297 (45.8%)
<b>TOTAL</b>	240 (100%)	269 (100%)	139 (100%)	648 (100%)

# Collaborators

---

## **Jen Dykema**

Survey Methodologist, UWSC

## **Kerryann DiLoreto**

Senior Project Director, UWSC

## **Dan Lawrence**

Programmer, UWSC

## **Nadia Assad**

Project Assistant, UWSC

Thank You!

For copies of this presentation or more information, contact:

Karen Jaques  
kjaques@ssc.wisc.edu

Please visit us at:  
**[www.uwsc.wisc.edu](http://www.uwsc.wisc.edu)**