Transcription reliability of hard-of-hearing and typically-developing preschoolers

Hannah Matalone[1], Casey Stoddard[2], Tracy Bookhout[2], Kellie Carns[1]

[1]Washington State University, [2]Eastern Washington University

hannah.matalone@email.wsu.edu

The Showcase for Undergraduate Research and Creative Activities

Pullman, WA | 30 March, 2015

Main research questions

1. **How do transcriptions vary between trained transcribers?**
2. **How do transcriptions of speech vary between typically developing and hard of hearing children?**
3. **Does transcription reliability vary according to increasing complexity of question type?**

Background

An enduring concern of developmental speech research is reliability of language transcriptions. This study looks at language samples collected from families with preschoolers with mild - to severe hearing loss (HH group) and a control group of families with typically-developing preschoolers (TD group).

By comparing how different question types are transcribed between these populations, we can determine if there is a high or low correlation between how similarly HH and TD recordings are understood.

Question Types

When transcribing speech, there are many different written variations of the same spoken word. For example, “momma” or “mama” can be used to correctly transcribe the same spoken word. In order to compare transcriptions with concrete variables, we compared the number of questions that were reported in each transcription. The different questions types are as follows listed in order of relative complexity:

1. **SAI – Subject Auxiliary Inversion**
2. **WHS – Wh-subject**
3. **WHA – Wh- adjunct**
4. **WHO – Wh-object**

Chart 1. Question hierarchy listed in order from least complex (1) to most complex (4).

Method

Participants

TD Families

29 families with a typically developing preschooler.

HH Families

14 families with a preschooler with mild-severe hearing loss.

Sample included boys and girls (mean age of ~30 months) who wore hearing aids and had no other disabilities. All children were involved in a larger longitudinal study.

Materials

Data was collected using the LENA system - (Language ENvironment Analysis; LENA foundation, Boulder, CO):

1. Digital Language Processor (DLP)

A small acoustic recording device which records up to 16 hours of raw audio on a solid state drive.

2. Automatic Speech Recognition (ASR) software

Custom software for analysis of *f*0 developed in MATLAB.

Procedure & Data Analysis

Using the LENA's automatic speech recognition (ASR) software and digital language processor (DLP), 15 minutes of high conversational activity was extracted from each of 43 whole-day recordings - 14 from HH families and 29 from TD families.

Two judges trained in phonetic analysis, Judge 1 and Judge 2, transcribed the same recording segments using CHAT in the CLAN program. Correlations among transcriptions were calculated to determine transcription reliability between these judges.

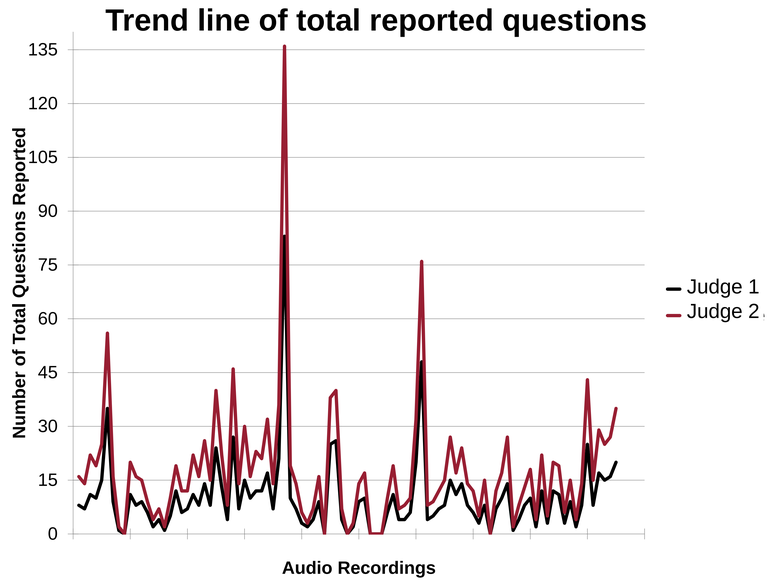

Figure 1. Total number of questions coded by two judges.

Question Type	SAI		WHS		WHA		WHO	
Hard of Hearing (HH) or Typically Developing (TD)	HH	TD	HH	TD	HH	TD	HH	TD
Correlation statistic	**r=.769**	**r=.823**	r=.213	r=.589	**r=.184**	**r=.223**	**r=.725**	**r=.954**
P-value	p<.0001	p<.0001	p<.0001	p<.0001	p<.0001	p<.0001	p<.0001	p<.0001

Table 1. Reported question correlation between Judge 1 and Judge 2 based on question type and hearing status.

Results

1. **The total number of questions reported by each transcriber are highly correlated.**
   a. Judge1 : Judge 2 Questions r=0.9968, p<.0001
2. **The scores given by Judge 1 and Judge 2 are highly correlated for the HH and TD population - TD showed slightly higher correlation.**
3. **For each question type (SAI, WHS, WHA, WHO) the reported quantity was correlated between Judge 1 and Judge 2. SAI and WHO are the most reliable overall.**

Conclusions

1. **The high correlation between the two transcribers demonstrates the reliability of the transcriptions and the clarity of the training manuals.**
2. **The high correlation between the transcribers for each question type/hearing status demonstrates the reliability in ability to transcribe and understand speech of all hearing types.**
3. **The relationship between syntactic complexity of questions and inter-rater reliability are fairly mixed. SAI and WHO questions have the highest correlation – they are most reliable. The correlation across the span of question difficulty suggests the high inter-rater reliability between the judges.**

Future directions

1. In this study, we can gain insight into how well the CHAT and CLAN manuals train users to operate this program and transcribe speech.
2. This research may serve as a foundation for research in the future requiring the use of speech transcription data or the CHAT and CLAN programs.