ChIMP: Children Interacting with Machines Project

1. Research Team

Project Leader:	Prof. Shrikanth Narayanan, *Electrical Engineering*
Other Faculty:	Prof. Elaine Andersen, *Neuroscience/Linguistics* Prof. Dani Byrd, *Linguistics* Prof. Margaret McLaughlin, *Annenberg School for Communication*
Graduate Students:	Alison Bryant, Simona Montanari, Serdar Yildirim
Undergraduate Students:	Sonia Khurana, Marni Landes, Talip Ucar

2. Statement of Project Goals

The long-term goal of this research program is to examine how children interact with machines in order to add naturalness and efficiency to spoken language interfaces. Children are comfortable and happy using spoken language interfaces, however systems must be tailored to understand child intent and provide a positive and successful experience. Performance of automatic speech recognition (ASR) systems degrades when the training and testing conditions are not similar. One reason for the acoustic mismatch between training and testing data is speaker variability. Typically, female speech is different from male speech. So do children differ from adults. Also, it has been shown that the effect of signal bandwidth on recognition performance is significant especially in dealing with children's speech. In order to design a more robust ASR system, it is important to know effects of ages and signal bandwidths on cepstral features that are widely used in ASR systems.

3. Project Role in Support of IMSC Strategic Plan

Communication and Education are two critical vision areas for IMSC. With limited fine motor skills and without the ability to read, write or type (well or at all), young children are one of the primary potential beneficiaries that use conversational interfaces, for example in games and computer instructional materials. Computer systems interacting with children need to be tailored for these users so that they will understand child intent and so that the child will have a positive and successful experience with the system. This research project supports that goal.

4. Discussion of Methodology Used

Developmental changes in the human speech production system signal age-dependent variability in the speech signal properties. ASR systems must accommodate the age-dependent characteristics of users within a developmental framework. Spectral and temporal variability in children’s speech is greater than that of adult’s speech. Also, children’s speech has higher pitch and formant frequencies, and longer segmental durations. This age dependence causes a serious degradation on ASR performance especially if, the ASR models are trained using speech from different age groups than those encountered during testing.

The researched examined developmental changes in the speech signal. The effects of age and signal bandwidth on speech signal features are analyzed especially motivated by implications to automatic recognition of children's speech. Mutual information is calculated between cepstral features and the vowel phonetic class for different age groups and signal bandwidths.

5. Short Description of Achievements in Previous Years

The results show that information contained in cepstral features about phonetic classes increases as bandwidth increases for all ages. Cepstral features of adult speech convey more information compared to that of children’s speech for both genders. Information increases rapidly between bandwidths 500Hz and 4500Hz. These findings based on mutual information correspond well with vowel recognition (classification) experiments. The vowel recognition experiment shows that as bandwidth increases recognition accuracy increases as well.

5a. Details of Accomplishments During the Past Year

Information contained in acoustic features about phonetic class for different ages and bandwidths is given in Figure 1 and Figure 2 for male and female speakers, respectively. It can be seen from the figures that information increases as bandwidth increases for all ages. It can easily be observed that cepstral features of adult speech convey more information compared to that of children’s speech for both genders. It should also be noted that between bandwidths 1500 and 3500 kHz, information difference between adult and lower age groups of male speakers are more significant compared to that of female speakers. Even there is an almost linear information increment as bandwidth increases across ages; Information contained in children speech cepstral features never reaches to adult level for both genders.

When results are examined with respect to ages, it is easily observed that cepstral features from adult speech convey much more information than that of children speech, ages 5 to 13 years, for both male and female speakers. Even though, there is significant information difference between children’s and adult’s speech cepstral features, one can not conclude that there is a linear increment between ages.

The vowel recognition results are given in Figure 3 and Figure 4 for male and female speakers respectively. It can be seen from the figures that accuracy rates increase as bandwidth increases. This can be explained by the (phonetic class dependent) information increase contained in the cepstral features as bandwidth increases. The accuracy rates degrade more rapidly for small ages compared to ages 16 and above, for both genders, as bandwidth decreases. It can also be concluded from these figures that recognizers that are trained with children speech are much more sensitive to bandwidth changes compared to recognizers that are trained with adult’s speech.

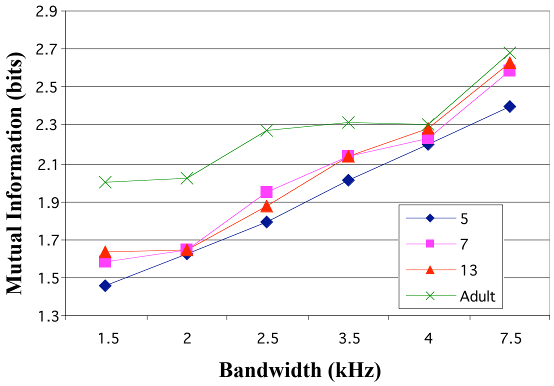

Figure 1. Information changes for different ages (years) with respect to different bandwidths (kHz) for male speakers.

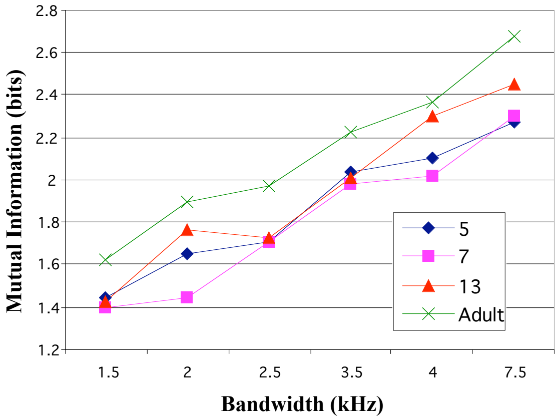

Figure 2. Information changes for different ages (years) with respect to different bandwidths (kHz) for female speakers.

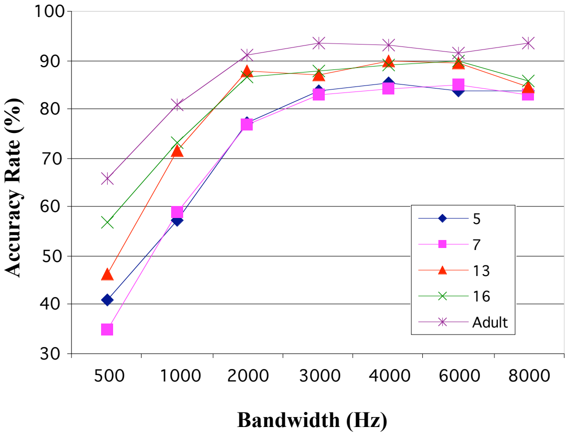

Figure 3. Vowel recognition accuracy results for male speakers.

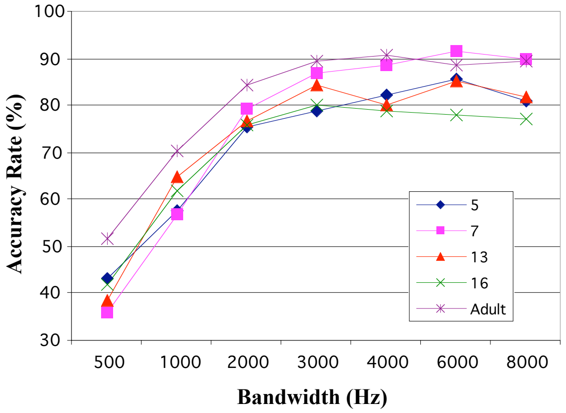

Figure 4. Vowel recognition accuracy results for female speakers.

6. Other Relevant Work Being Conducted and How this Project is Different

A detailed description of the related work can be found in http://sail.usc.edu/chimp

7. Plan for the next year

- Analysis of acoustic characteristics of preschool children
- Analysis of gesture behavior in child-computer conversation.

8. Expected Milestones and Deliverables

- Improved ASR capabilities for children.
- Contributions to interactive educational systems.

9. Member Company Benefits

N/A

10. References

[1] S. Narayanan and A. Potamianos, “Creating conversational interfaces for children,”" *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 2, pp. 65-78, 2002.

[2] S. Lee, A. Potamianos, and S. Narayanan, “Acoustics of children's speech: Developmental changes of temporal and spectral parameters,” J. Acoust. Soc. Am., vol. 105, pp. 1455-1468, Mar. 1999.

[3] Miller, J. D., Lee, S., Uchanski, R. M., Heidbreder, A. H., Richman, B. B., and Tadlock, J., “Creation of two children’s speech database,” in Proc. of ICASSP, (Atlanta, GA), 1996.

[4] Li, Quan and Russell, Martin J., “Why is Automatic Recognition of Children’s Speech Difficult?”, Eurospeech 2001, Scandinavia.

[5] T.M. Cover and J. A. Thomas, “Elements of Information Theory,”(Wiley Series in Communications), John Wiley & Sons Inc., New York, 1991.

[6] Li, Quan and Russell, Martin, "An Analysis of the Causes of Increased Error rates in Children’s speech Recognition,” in Proc. of ICSLP, (Denver, CO), 2002.

[7] Padmanabhan, M., “Use of Spectral Peak Information in Speech Recognition”, Speech Transcription Workshop, NIST, University of Maryland, 2000.

[8] Young S. J, Odell J., Ollason D., Valtchev V., Woodland P., “HTK Book” Cambridge Research Laboratory, 1997.