From HSR

AuditoryModels: Auditory Models

A collection of software, research, history, reflections, and data related to auditory models.

Demos and Software

Talk at Johns Hopkins on human phoneme recognition

The Role of the Cochlea in Human Speech Recognition from CLSP Seminars PRO Jont Allen, UIUC 2007 August 7 Center for Speech and Language Processing, Johns Hopkins University **vimeo video**

Harvey Fletcher mp3 video (~28 mins long)

Harvey Fletcher movie from c1963

Allen Video about modeling cochlear transduction (ca. 1997)

Videos

Interspeech 2013 Tutorial presentation

- Interspeech-2013 Lyon France Tutorial T4:
    1. Tutorial **Index of topics**
    2. Part I **pdf**
    3. Part II **pdf**
    4. Part III **Demos**

Interspeech 2013 Demos of *Cue-modified* speech

- **da -> ga**
- **ta remove burst**
- **ka -> ta**
- **Sa -> sa**
- **Sa -> da**
- **Sa 2 sa 2 cha 2 za 2 Da**
- **za voicebar** The very significant voice-bar plays no role in perception of

za.
- **Sa vs. Za** What is the cue (Not duration)? For the answer read this **pdf**
  1. Interspeech 2013 Demo files: **wav, mv4**
  2. Interspeech 2013: All-files (talks, wav, video, demos) **tgz**

Demos (These older demos are inferior to the Interspeech-2013 Tutorial demos, above)

- *KunLun* software to analyze and modify speech (wav format), *using the AI-gram* software **KunLun (zip)** and wav files **example phrases (zip)**
- Demos of what *KunLun* can do **Video-demos**
- Support documentation that describes the basic speech perception research behind KunLun:
  1. Allen, Jont and Li, Feipeng (2009). *Speech perception and cochlear signal processing*, *IEEE Signal Processing Magazine*, Invited: *Life-sciences*, **26**(4), pp 73-77, July. (**pdf**, **djvu**)
  2. Feipeng Li and Jont B. Allen. (2011) Manipulation of Consonants in Natural Speech; IEEE Trans. Audio, Speech and Language processing, (officially published: Jul, 2010; Appearance date: Mar 2011) pp. 496-504 (**pdf**)
  3. Li, F., Menon, A. and Allen, Jont B., (2010) A psychoacoustic method to find the perceptual cues of stop consonants in natural speech, apr, J. Acoust. Soc. Am. pp. 2599-2610, (pdf)
  4. Li, F., Trevino, A., Menon, A. and Allen, Jont B (2012). "A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise" J. Acoust. Soc. Am., v132(4) Oct, pp. 2663-2675 (pdf)
- AIgram source code **zip, txt**; If you would like to download this code, ask me for the password.

Research Objectives and Accomplishments

The research in the Human Speech Recognition group is directed at a fundamental understanding of speech perception in both normal-hearing (NH) and Hearing-Impaired ears. These are related problems, and are actually a continiuium, not two separate things. Most people are born with normal hearing. Within a few years we learn, without seeming effort, to understand human speech. How this happens is a mystery. But what happens is not a mystery. The research we have been doing over the past 10 years, as documented in the section below, is a systematic study of the nature of the failure to process and communicate under various conditions. Only by stressing the system, causing failure, can we hope to understand it. There are at least four levels of experimentation:

1. The first level of experiments is with NH ears, with speech in noise.
2. The second level of experiments are filtering experiments, where the speech is filtered before the noise is added.
3. In the third series of experiments, the speech is truncated in time.
4. Finally small regions of speech are modified by a few dB, or removed altogether.

Examples of such processing are given in later on this page.

Findings:

We have found that speech perception is a discrete (binary) zero error task **Singh and Allen, 2012**. Working at the token level, we defined 2 groups: ZE, NZE. *Zero-Error* (ZE) speech is defined as speech that NH listeners *never* make an error in identifying, at and above above -2 dB SNR. The non-ZE (NZE) sounds are all the rest. All of the speech CV sounds that we have tested contain many ZE tokens: most CV consonants consist of more than 80% ZE utterances.

The remaining 20% of the CVs may be broken down into 0% < medium-error (ME) <10% and >10% high-error (HE) groups. ME consonants are typically utterances having varying degrees of mispronounced utterances. HE consonants are typically those that are heard as a different sound, with high probability (>20%). Based on the entropy across normal hearing listeners, we view such sounds as *mislabled.* The reasons for these errors can typically be traced to a specific flaw in the production of the sound, which is typically easily identified.

A chronological history of HSR papers

Summary of UIUC-HSR Experiments (Updated Mar 15, 2014)

Year	Experiment	Students	Details; N_s=# Subjects	Publications	.mat
2004	MN64 (MN04SWN)	Phatak & Lovitt	Miller-Nicely in SWN with 4 vowels: f/a/ther, b/a/t, b/i/t, b/ee/t (not b/e/t) i.e., LaTex's tipia ``textipa{ @, \ae, E, i}, *LDCbet: [a, xq, i, xi] ([a, Q, i I]),*	Phatak & Allen (2007) [PA07] **pdf**	**MN64**

			V_{ldc}=/a, @, i, I/ $N\_s = 4 - 4$bad subjects''		
2005	Study	Allen, J. B.	Consonant recognition and the AI	JASA **117**(4), p. 2212-2223. (2005) **pdf**	
2005	MN16-R (MN05WN)	Phatak & Lovitt	Replicate MN04 (WN)	Phatak, Lovitt & Allen (2008) **pdf**	
2005	MN64R (MN05SWN)	Phatak & Lovitt	More MN64; **14 new subjects**; SWN	**pdf**	**MN64**
2005	HIMCL05	Yoon & Phatak	CVs; 10 HI ears @MCL in WN	Phatak, Yoon, Gooler & Allen (2009) **pdf**	
2006	HINALR05	Yoon	CVs; 10 HI ears; NALR@MCL in SWN		
2006	Verification	Regnier	Modifications of /ta/	Regnier & Allen (2008) **pdf**	
2006	CV06SWN	Phatak	C_{ldc}=/d,b,k,p,s,t,S,Z,z/, V_{ldc}=/o,E,u,R,Q,U,I,a/		**cv06swn**
2006	CV06WN	Regnier	9C+8V WN /d, b, k, p, s, t, xs, xz, z/		**cv06wn**
2007	CV06	Pan	Analysis of 9 Vowels of CV06	2 unpublished MSs	
2007	HL07	Li	High and Low pass Repeat of Fletcher	Li Allen 2009, JASA **pdf**	
2008	TR07	Li	Time Truncation after Furui86	Allen Li (2009) ASSP Magazine **pdf**	
2008	TR08	Li	Time Truncation after Furui86	? 3 vowels ?	
2009	3DDS	Li	3DDS (i.e., MN64, HL07, TR07-8)	Li Allen (2010) JASA; Li Allen (2010) IEEE TLSP;	

				Li Trevino Allen 2012 JASA	
2009	Verification	Menon	Remove Primary burst		
2009	Verification	Abhinauv	Modify (± 6 dB)+Remove Primary burst	Kapoor and Allen, 131(1), 2012 **pdf**	
2009	Verification	Cvengros	Modify burst + devoiced + voiced transition		
2009	MN64(+R)	Singh	Full analysis of N_s =25 of MN64+MN64R	JASA, April 2012 **pdf**	
2010	HIMCL10-I/III	Woojae Han	CVs; N_s=46 HI ears with N_t=2/token/SNR	**pdf**	
2010	HI10NALR-II/IV	Woojae Han	CVs N_s=17 HI ears with N_t=10/token/SNR	**pdf**	
2011	HL11	Trevino	High/Low filter CVs of HI10		
2013	HI Exp2 Analysis	Trevino	Analysis of the individual variability of HI	Trevino & Allen **pdf**, **pdf**	

Databases

- LDC isolated **CV, VC, CVC, VCV** sounds
- LDC **phrases** that work with KunLun
- LDC symbol definitions *LDC_symbols*.(**tex**,**pdf**).
- Cross-check by vowel synthesis: **IPA-With-Sound**
- Unicode **database** vs. **Microsoft fonts**

Software of interest

- To get past matlab issues, this may be the future: **Python matlab wrapper**

Measurement systems

- **ARTA** software to be used with your sound card. Performance will vary
- **QA400** Inexpensive (e.g., $\{\backslash tiny \approx \}$\$200) USB box with Windows software with a -140 [dB] noise floor and 110 [dB] dymanic range.

HSR Pictures (entertainment value only)

- IHCON 2010: presentation **click** $^\Delta$, hiking with Jont Allen, Brian Moore, Stefan Launer, Woojae Han, Riya Singh and Angali Menon **jpg** $^\Delta$, and biking **jpg** $^\Delta$
- Mead Killion visits HSR (5/7/2010): **jpg**
- **ICSLP 2006 jpg**
- **ASRU 2009**
- Parties for **Bob Shannon 2004** and Chris Shera
- Third Mechanics of Hearing (Kemp and Brown, 1988) (Historic **Photos**) Keele England

Historical Documents

- Interesting views: **Chomsky Lectures**
- **IEEE History of Hearing Aids**
- **Historical Books**
- *Speech and Hearing* by Harvey Fletcher **djvu**
- Tables from Miller & Nicely 1955, "An Analysis of Perceptual Confusions Among Some English Consonants" **Attach:mn55.zip** $^\Delta$

Retrieved from http://173.161.115.243/pmwiki.php/AuditoryModels/HomePage
Page last modified on June 22, 2015, at 08:24 AM