"Getting Started with the CU-Move Corpus"

http://cumove.colorado.edu/

John H.L. Hansen
Center for Spoken Language Research
(3215 Marine St. for express mail)
University of Colorado Boulder
Boulder, Colorado 80309-0594 U.S.A.
http://cslr.colorado.edu/
email: John.Hansen@colorado.edu
Tel: (303) 735-5148, Fax: (303) 735-5072

Release 2.0B (March 4, 2005)
Release 2.0A (November 17, 2002)

In this report, we provide information for individuals interested in using the CU-Move speech corpus. This in-vehicle corpus consists of the largest collection of speech data available for in-vehicle route navigation and planning with a wide range of noise conditions and speakers from across the United States. This report will cover the following six areas.

Outline

Release 2.0B includes a number of conference and journal papers which have used the CU-Move corpus. We will continue to include copies of papers from other groups in order to share results and resources.

Section 1.0: Overview of CU-Move Corpus

The goal of the University of Colorado CU-Move project is to develop algorithms and technology for robust access to information via spoken dialog systems in mobile, hands free environments. The novel aspects include the formulation of a new microphone array and multi-channel noise suppression front-end, corpus development for speech and acoustic vehicle conditions, environmental classification for changing in-vehicle noise conditions, and a back-end dialog navigation information retrieval sub-system connected to the WWW. While previous attempts at in-vehicle speech systems have generally focused on isolated command words to set radio frequencies, temperature control, etc.,the CU-Move system is focused on natural conversational interaction between the user and in-vehicle system. Since previous studies in speech recognition have shown significant loses in performance when speakers are under task or emotional stress, it is important to develop conversational systems that minimize operator stress for the driver. System advances include intelligent microphone arrays, auditory and speaker based constrained speech enhancement methods, environmental noise characterization, and speech recognizer model adaptation methods for changing acoustic conditions in the car. Our initial prototype system allows users to get driving directions for the Boulder area via a hands free cell phone, while driving in a car.

As part of CSLR's commitment to advancing the state-of-the-art in conversational dialogue systems for in-vehicle route navigation, CSLR undertook the task of establishing and organizing a two phase in-vehicle collection plan (details of this plan can be found on our web site: http://cumove.colorado.edu/). The site contains more information on the collection plan, example prompts, sample audio clips, consent forms, etc.

- Phase I: Acoustic Noise Data Collection and Analysis
- Phase II: Speech & Speaker Data Collection
    - Collection plan, examples of prompts, sample audio data
    - Human Subject Consent Form
    - Human Subject Information Form
    - Complete CU-Move set of prompts : (CU-MovePrompts.zip)

The CU-Move data consists of three hard disk releases:

- **Hard-Disk#1:** CU-Move Release 1.1A: {60GB disk size}
    - Documentation File [CU-Move-CorpusRel-1.1a.Jan02]
    - Minneapolis, MN: all 153 speakers collected from Minn.,MN
    - 153 speaker directories (labeled 0050 through 0205)
- **Hard-Disk#2:** CU-Move Release 2.0A: {80GB disk size}
    - Documentation File [CU-Move-CorpusRel-2.0A.Nov02]
    - St. Louis, MO: all 152 speakers collected from St. Louis, MO
    - Manchester, NH: all 124 speakers collected from Manchester, NH
- **Hard-Disk#3:** {40GB disk size}
    - Savannah, GA: 32 speakers
    - Dallas, TX: 30 speakers
    - Boulder, CO: 25 speakers

[suggested disk: Western Digital Caviar, 7200rpm, IDE, 80GB or 100GB]

This write-up consists of all necessary documentation which was contained in Release 1.1A (note that Hard-Disk#3 release will not contain any further documentation than what is contained in this document). This document contains a more detailed description of the collection protocol, speaker population profiles (age, education, gender, etc.), transcription details, environmental noise analysis, speaker training and testing lists.

This document is intended to provide the necessary description of how data was processed and organized on your hard disk distribution.

Section 2: Directory Structure for Hard Disk Distribution

Disk #1:	Minneapolis, MN:	Speakers	0050 -- 0205
Disk #2:	St. Louis, MO:	Speakers	0206 -- 0360
	Manchester, NH:	Speakers	0361 -- 0485
Disk #3:	Savannah, GA:	Speakers	0486 -- 0517
	Dallas, TX:	Speakers	0519 -- 0548
	Boulder, CO:	Speakers	0001 -- 0028

As an example from Disk #1: Each speaker directory is numbered 0050 through 0205. Each speaker directory contains the following files or sub-directories:

0050-03.sph
0050-06.sph
0050-07.sph
0050-BF.sph
0050-03.trs
/Navigation
/Digits
/Streets
/Sentences
/Dialog

All speech files are in NIST Sphere header format (contact NIST for information, we used their Sphere2.6a package). Each audio segment is 45 minutes in duration, and composed of 5 areas broken into Part 1 and 2 (summarized below). Each of the sub-directories above are empty. We provide a Perl script and C-code to automatically extract each audio segment area (i.e., Digits, Streets, etc.).

PART 1: *Structured Text Prompts*. The driver performs a fixed route similar in structure to what was done for Phase 1 data collection that includes a combination of driving conditions (city, highway, traffic noise, etc.) for each speaker. Prompts were given from a laptop display situated around the glove compartment of the vehicle. This portion takes 30 minutes to complete. There are four subsections that include:

[1.] NAVIGATION Direction Phrases section: a collection of phrases which are determined to be useful for In-Vehicle navigation interaction [prompts fixed for all speakers]
[2.] DIGITS prompts section: strings of digits for the speaker to say [prompts randomized]
[3.] STREETS / Address / Route locations section: street names or locations within the city; some street names will be spelled, some just spoken. [prompts randomized]

[4.] SENTENCES - General Phonetically Balanced Sentences section: collection of phonetically balanced sentences for the speaker to produce [prompts randomized]

PART 2: DIALOG Wizard - of - Oz Collection. Here, the user calls a human "wizard" (WOZ) who guides the subject through various routes determined for that city. More than 100 route scenarios particular to each city were generated so that users would be traveling to locations of interest for that city. The human WOZ had access to a list of establishments for that city where subjects would request route information (e.g., "How do I get to the closest police station?" "How do I get to the Hello Deli?"). The user would call in with a modified cell-phone in the car, that allows for data collection using one of the digital channels from our recorder.

Data Preparation:

Each of the 44kHz audio streams must be processed and organized for distribution. The sequence of data preparation is:

1) Channels 1-5 were submitted to the Beamformer code to obtain 1 beamformed channel output which is still at the 44kHz sample rate.

2) The following channels were then submitted to the downsample code to produce 16kHz output signals: (files from speaker 0050 shown alongside)
   a. 44kHz Beamformed output ***0050-BF.sph***
   b. Channel 3 (center channel of array) ***0050-03.sph***
   c. Ref. Channel ***0050-06.sph***
   d. Cell Phone mike (if available) ***0050-07.sph***
   e. AKG mike (if available) ***0050-08.sph***

   Again, the outputs here were stored in uncompressed SPHERE header format.

Notes:

- For RELEASE 1.0A (Minneapolis, MN) data, the AKG microphone was not available, so channel spkr_08.sph is not present for any of the speakers. Also, for a small set of the 153 speakers, the Cell Phone channel (spkr_07.sph) was not available (this was due usually to a case when the human WOZ was not available at the route server back at CSLR to engage the subject in route navigation).
- For RELEASE 2.0A
  - St. Louis, MO: this data contained the AKG microphone (powered by a power supply built at CSLR). This city contained WOZ and AKG for all speakers (some exceptions noted in the speaker summary sheets).
  - Manchester, NH: this data contained WOZ dialog portion, some AKG microphone recordings.
  - Savannah, GA; Dallas, TX; Boulder, CO: this data contains array and reference microphone recordings. No AKG or WOZ data is available for these cities (in an effort to reduce costs, WOZ support was dropped after Manchester, NH).

Note: in Sec. 5, detailed speaker information as well as identification of which channels are available for each speaker are provided.

Data Distribution Format:

Since a large amount of data has been collected (over 700 GB), we are distributing CU-Move data via hard disk. Center members interested in receiving the data should purchase their own hard disk **(suggested disk: Western Digital Caviar, 7200rpm, IDE, 80GB or 100GB**), [see Pg. 2 for summary of disk size for the 3 releases], and forward it to CSLR. We will download speech data and transcriptions and return the disk. As further portions of the estimated 700GB of data have become available (transcribed, downsampled/organized for distribution), we have distributed the corpus via hard disk (we fill up your hard disk and return).

The disk contained in this shipment was formatted on a Linux machine [we are presently using FAT32, so the disk can be read from a Linux or Windows machine]. The procedure for extracting data is as follows:

1) Create a mount-point directory on your Linux machine (something like /home/cumove)
2) Install the hard disk into your machine (we installed as a “slave” drive, with the assumption that your main drive is the “master” which has the operating system loaded on it).
3) Execute the mount command:

   Mount -t ext2 /dev/hd{a,b,c,d} /home/cumove

   Note that normally hard disk “a” is your master with the operating system, so most likely the drive will be either {b,c,b}, depending on how many drives you have already on your machine. After mounting, you should be able to access all the data on the hard disk.

CU-Move Cell-Phone Data (part of release 1.1A):

The CU-Move Wizard-of-Oz Dialog component had a modified cell phone which allowed for direct recording onto the Fostex unit. There was also audio collected back at CSLR for the human WOZ who was talking with the subject in the field. There is a Cell-Phone directory on the disk that is provided “as is”, but would provide useful information on telephone channel issues during WOZ dialog collection. There may not be perfect alignment with transcripts provided with Channel 3 (the transcriber tool time stamps a begin point before speech is provided). The data is organized as follows:

0050-001mn	speaker 50, using route scenario 1 from Minn., MN
0050-002mn	speaker 50, using route scenario 2 from Minn., MN
0058-015mn	speaker 58, using route scenario 15 from Minn., MN
0058-RTEmn	speaker 58, describing actual driving route back to pickup

These represent the entire audio stream collected at CSLR (from the WOZ side after passing through the cell phone in the field and telephone network). Example 0050-001mn is speaker number 50, “001” means that route scenario “1” was selected (about 100 unique routes for each city were formed and are included on the hard disk as well under WOZ-routes). This provides details on street names for directions. The last format file in this directory is the 0058-RTEmn file. Any “RTE” files contain subjects describing what they see as they are driving back to the base pickup point (what they see out the window). This represents the most natural dialog speech and occurs while the car is moving to a destination (the other route scenarios focus on the subject discussing route plans, but the vehicle is not actually moving to that location because of time collection restrictions).

Section 3: Transcription Details & Results

In this section, we discuss CSLR's transcription work on the CU-Move corpus. Since a number of transcribers participated, we also performed a spell check analysis and an evaluation of inter-labeler reliability.

We used Channel 3 from the microphone array for all transcription efforts. The focus was on obtaining text transcriptions of all data produced in the 45minute stream. The files needed for transcription are as follows:

1) Produce a transcription file for EACH of speaker, with a text segment identifier that indicates each of the 5 CU-Move audio Parts (i.e., within the transcription file 0050-03.trs, there are 5 text identifiers that say: Part 1, Part 2, …, Part 5, that identify the starting location where each of the 5 data collection parts begin (Navigation, Digits, Streets, Sentences, Dialog).
2) All transcription was done with the LDC Transcriber tool (see the LDC web page for a free download and documentation of the transcription tool).
3) Transcription of Part 1: *Structured Text Prompts*, focused on verification that the speakers produced what was prompted for on the dashboard display. Each speaker would advance the display using a handheld wireless mouse. The four parts of this area included: NAVIGATION Direction Phrases, DIGITS, STREETS / Address / Route locations section, and SENTENCES - General Phonetically Balanced Sentences.
4) Transcription of Part 2: *DIALOG Wizard - of - Oz Collection*, focused on providing word/sentence level transcription of the spontaneous dialog between the field subject and the CSLR WOZ operator for each of the field subject's selected points of route navigation.

A group of three transcribers worked to perform verification and word level transcription of the CU-Move corpus. After transcription work was completed, a spell checking task was performed to verify consistency across transcribers for spelling and labeling.

Finally, in an effort to assess the inter-labeler reliability across the three transcribers, we performed a blind transcription task where each of the transcribers independently and without their knowledge, transcribed a small set of CU-Move speakers. Next, using these transcription files, we performed a test using the NIST sclite code to determine differences in output transcripts. The table below summarizes a comparison based on Parts 1-4 (which consisted of prompted speech, and therefore should have higher labeler agreement), along with Part 5 (which consists of WOZ spontaneous speech, and therefore would possess more variability between transcribers). Clearly the performance numbers will depend heavily on which speakers are selected (if one subject speaks softly and mumbles their voice, there will be higher disagreement between transcribers; versus a speaker who clearly articulates their speech). We should examples from two speakers below in Tables 1 and 2.

TRANSCRIBER **Ref#-to-Hypoth#**	PERFORMANCE [Part 1-4]			
	DEL	SUB	INS	WER
TRANS: #1-to-#2	2.7	1.6	5.8	10.1
TRANS: #1-to-#3	5.6	1.8	5.1	12.5
TRANS: #2-to-#1	5.7	1.5	2.6	9.8
TRANS: #2-to-#3	4.4	2.0	0.8	7.1
TRANS: #3-to-#1	5.1	1.8	5.7	12.5
TRANS: #3-to-#2	0.8	2.0	4.6	7.4
AVERAGE	**4.0**	**1.8**	**4.1**	**9.9%**

Table 1: Transcriber agreement between 3 CU-Move transcribers using Parts 1-4 for sample utterance 002 (i.e., approximately 30 minutes of speech material, majority of speech was prompted with an in-dash display.)

TRANSCRIBER **Ref#-to-Hypoth#**	PERFORMANCE [Part 5]			
	DEL	SUB	INS	WER
TRANS: #1-to-#2	6.9	3.1	16.	26
TRANS: #1-to-#3	14.6	2.9	6.3	23.8
AVERAGE	**10.8**	**3.0**	**11**	**25%**

Table 2: Transcriber agreement between CU-Move transcribers using Part 5 (WOZ) for sample utterance 002 (i.e., approximately 15 minutes of speech material, majority of speech was spontaneous navigation dialog material.)

Table 1 shows that across the three transcribers, there was significant agreement in transcription text and notation for Parts 1-4 (Navigation, Digits, Streets, Sentences). The average WER was 9.9%. We should point out that this will vary across the speakers in the corpus, and since it takes approximately 6 hours to transcribe one speaker, we only tested a small set for the three transcribers. Table 2 shows transcriber agreement for Part 5, the human-to-human WOZ route navigation dialog portion. As expected, since this is spontaneous speech, there is much more variability in how transcribers mark this material. We see that there is an average 25% WER across the two transcribers (we did not test the third transcriber, since he performed less of the WOZ transcription effort).

The results from Tables 1 and 2 suggest that strong agreement exists between transcribers for the +500 CU-Move speaker corpus. We should again point out that these numbers will vary across individual speakers depending on their regional accent and articulation traits.

Section 4: <u>In-Vehicle Environmental Noise Analysis</u>

One of the primary goals of the CU-Move corpus is to collect speech data within realistic automobile driving conditions for route navigation and planning. Prior to selection of the vehicle used for Phase II data collection across the United States, and in depth acoustic analysis was first performed on six vehicles in Boulder, Colorado. This section briefly summarizes this analysis and the noise analysis findings.

Vehicles:

A set of six vehicles were selected for in-vehicle noise analysis. These vehicles were model years of 2000 or 2001 (all had odometer mileage readings which ranged between 11 – 8,000 miles). The six vehicles were:

- [Cav] Chevy Cavalier Compact Car
- [Ven] Chevy Ventura Mini-Van
- [SUV] Chevy SUV Blazer
- [S10] Chevy S10 Extended Pickup Truck
- [Sil] Chevy Silverado Pickup Truck
- [Exp] Chevy Express Cargo Van

The Cav, Ven, and SUV were leased from a local rental car company, and all had mileage odometer readings between 3000-8000 miles. Since pickup trucks and cargo vans were not easily obtained from local rental car companies, arrangements were made to obtain access to the remaining three vehicles from a local GM car dealer. The [S10] Extended Cab pickup truck had 9 miles total on its odometer. The Silverado [Sil] pickup truck had a total of 17 miles on its odometer, and the Chevy Express [Exp] cargo van had an odometer reading of 21 miles. All vehicles were 2001 models. Figure 1 shows sample images of two of the six vehicles that were used for acoustic noise data collection.

Figure 1: Images of vehicles used for acoustic noise data collection
(Chevy Express Van and Chevy S10 Extended Cab pickup truck)

Recording Setup:

Having selected the vehicles, we now turn to the data collection setup. A five channel microphone array was developed (designed, constructed, and tested) for use in interactive speech systems by researchers at the Center for Spoken Language Research (CSRL) under support from DARPA. Figure 2 shows an image of the five-channel microphone array (three additional arrays were constructed using a designed PC board from the proto-type shown, and used for Phase II collection across the United States). The array was constructed using Knowles microphones. The microphone array was evaluated on a per channel basis for power supply and pre-amplification transfer function. Individual channel outputs were directed out from the array for each microphone (i.e., individual microphone channel outputs were wired to a multi-unit connector at one end of the array). In addition to this array, a reference noise microphone and housing was also developed and constructed by CSLR. Figure 2 shows the CSLR microphone array and digital recorder. The following discussion considers the microphone array frequency response and microphone pre-amp circuit. Figures 6,7, and 8 show figures of four of the six vehicles where microphone array, reference microphone, and digital channel recorder setup is shown. For all vehicles, the data recording unit was powered from an auxiliary power source from each vehicle. A DC-to-AC power converter was used to supply power to the Fostex 8-channel digital recorder, Shure mixer which was modified to work as a pre-amplifier, and a constructed DC power supply for the reference and microphone array. For all data collection, the engine was started and allowed to run for several minutes before powering up the supply on the recorder unit (i.e., the Fostex, then pre-amp, then microphone power). This allows the higher startup surge power for the Fostex unit to settle before power is drawn for the preamp and microphone power supply. During all recording sessions, the recording power system worked without any problems (no fuse problems, or high current drain during startup).

Figure 2: CSLR Microphone array (early proto-type) and constructed 8-channel digital recorder system.

Microphone Characteristics for CU-Move Data Acquisition:

The 5-channel microphone array and reference microphone used for CU-Move data collection was designed, built, and tested at the CSLR. Both input devices use the Knowles EK3024 12S microphone. This microphone has an output impedance of 2.8kOhm to 6.8kOhm at 1kHz, with a <10V supply, and a sensitivity of -53 dB re (1Vrms/0.1Pa). The response curve is shown in Figure 3. This plot is for an open circuit condition with a 1.3V supply.

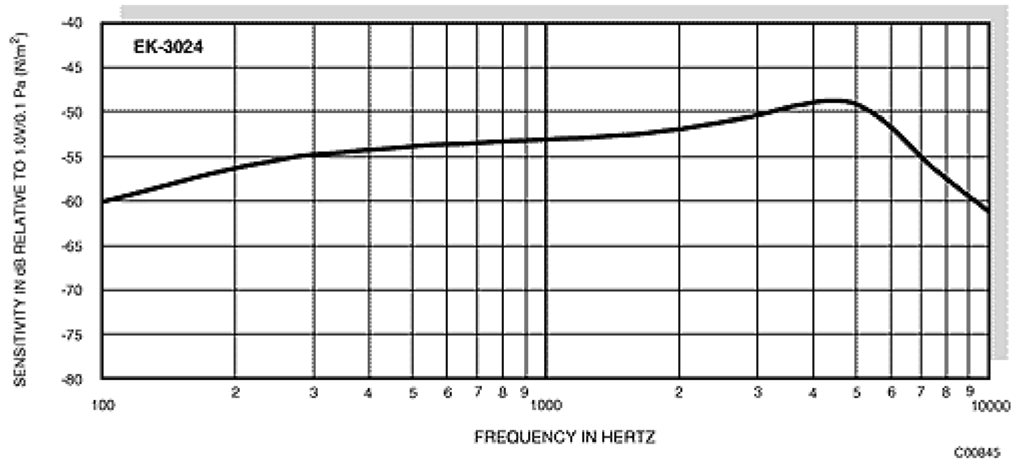


Figure 3: Knowles EK3024 12S microphone frequency response used for array and ref. mike.

Microphone Gain Circuit: The microphone filtered gain circuit used for the individual microphones in the array and reference microphone is shown in Figure 4. It is composed of a Multiple Feed Back third order band pass configuration with a Butterworth response, and a single ended to differential output. The gain of this circuit is controlled by the resistor 2x kOhm.

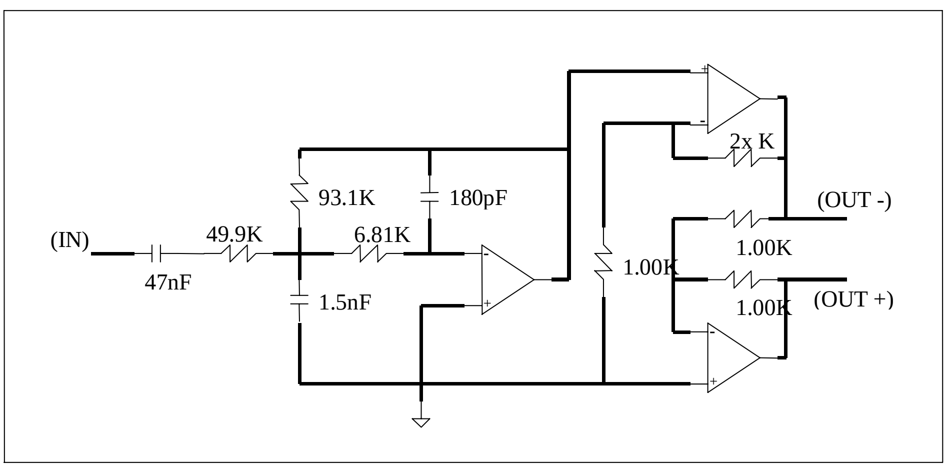


Figure 4. Microphone Filtered Gain Circuit.

Note that the response of this circuit takes into account the output impedance of the microphone as a source impedance for the node marked (IN). Thus, the effective impedance in the branch with the 47nF capacitor is, in the worst case, 6.8kOhm + 49.9kOhm. The frequency response of the microphone filtered gain circuit, approximated for 500hZ and above, can be expressed as

$$\frac{v_o}{v_i} = -\left(R2/R1\right)\frac{1/R2R3C1C2}{s^2 + 1/C1\left(1/R1 + 1/R2 + 1/R3\right)s + 1/R2R3C1C2}$$

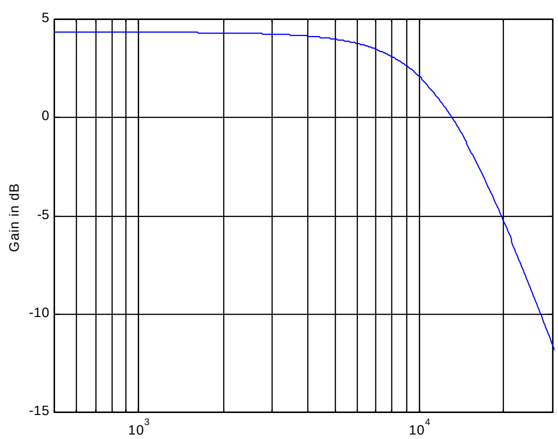
$$\frac{v_o}{v_i} = -(1.65)\frac{5.84e9}{s^2 + 1.17e5s + 5.84e9}.$$

The response is plotted in Fig. 5a. The -3dB point and gain taken from the plot are 12kHz and 4dB respectively. The frequency response below 500Hz can be expressed as,

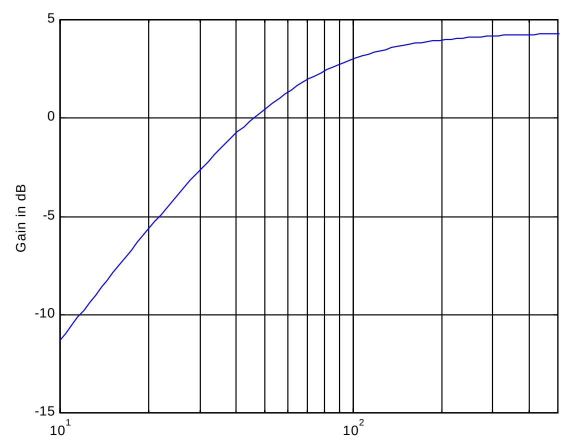
$$\frac{v_o}{v_i} = -\left(R2/R1\right)\frac{s}{s + 1/R1C}$$

$$\frac{v_o}{v_i} = -(1.65)\frac{s}{s + 377}.$$

The response is plotted in Fig. 5b. The -3dB point and gain taken from the plot are 60Hz and 4dB respectively. The fixed gain of this response excludes the gain set by the x kOhm resistor in Figure 19. The overall gain for the array and reference microphones is empirically set to –50dB (re 20uP) and can be modified by adjustment of resistor 2x kOhm. Given the EK 3024 sensitivity of 1.0V/0.1Pa, assumed at 1kHz, the overall sensitivity of the array and reference microphones is expected to be 159V/Pa.



a) Approximation f > 500 Hz



b) Approximation f < 500 Hz

Figure 5. Amplitude Frequency Response of the Array and Reference Microphones.

Figure 6: <u>Chevy Express Van:</u> Data Collection Setup: Microphone array position, digital data recorder position, Reference microphone and positioning behind driver seat.

Figure 7: <u>Chevy S10 Extended Cab Pickup Truck:</u> Data Collection Setup: (a) Microphone array position, (b) digital data recorder position, (c) Reference microphone and positioning behind driver seat (note: photo taken through rear window towards driver's seat).

Figure 8: <u>Chevy Silverado Pickup Truck:</u> Data Collection Setup: Microphone array position, digital data recorder position, Reference microphone and positioning behind driver seat (note: photo taken through rear window towards driver's seat).

Figure 9: <u>Chevy SUV Blazer:</u> Data Collection Setup: Microphone array position, digital data recorder position, Reference microphone and positioning behind driver seat.

Route and Vehicle Setup:

Since the focus of our Phase I evaluation was to determine a vehicle that would broadly represent vehicles for in-vehicle route navigation and information access, a route was planned which contained a sample of all driving conditions expected for use in city and rural areas. Figure 10 shows the approximately 18 mile route which began at Scott Carpenter Park [Boulder, CO] (which was used as a staging area, since we could test recording equipment for sound levels with limited outside traffic noise). Stop lights are indicated at each intersection encountered, as well as locations where a predetermined phonetically balanced sentence was produced. Note that the microphone array and reference microphone were positioned on the driver side of each vehicle. Data collected in Phase II had the microphone array on the passenger side. Recording began at the PARK, with windows closed, AC off. The route up until Sentence #1 represented moderate traffic (several stop lights, speeds approximately 40-45 mph). At Sentence #1, this is after acceleration onto the Boulder Turnpike (Route 36) where the speed limit is 65mph (4 lane divided highway). We traveled south-east on Rt-36 until we took an exit ramp near a Costco store which is the Louisville-Superior exit [about 5 miles from Boulder] (this was an exit ramp with an overpass, 2 stop lights needed until we could go back onto Rt-36 traveling back north-west towards Boulder again). Upon entering Rt-36 again, the passenger and driver's windows were lowered by about 2 inches. Upon reaching Colorado Ave., we turned into Univ. Colorado Boulder campus, and drove up through Broadway and finally onto Arapahoe Ave. This area is very high traffic area with many students crossing the 4-lane road, bicycles, stops lights – (stop and go traffic), etc. Speeds may have reached 30mph, but only for short periods of time. Windows were rolled down ½ way for both the driver and passenger side windows when we reached Arapahoe Ave. When we reached the intersection of 28th St., the windows were rolled up and the AC turn on with fan at full speed. This was kept on until we reached the PARK on 30th St., at which time the AC was turned off; and the wiper blades were turned on, wiper fluid used with wipers on, and wiper blades turned off. The production of phonetically balanced TIMIT sentences was used to represent anchors along the route for ease in labeling the acoustic recordings.

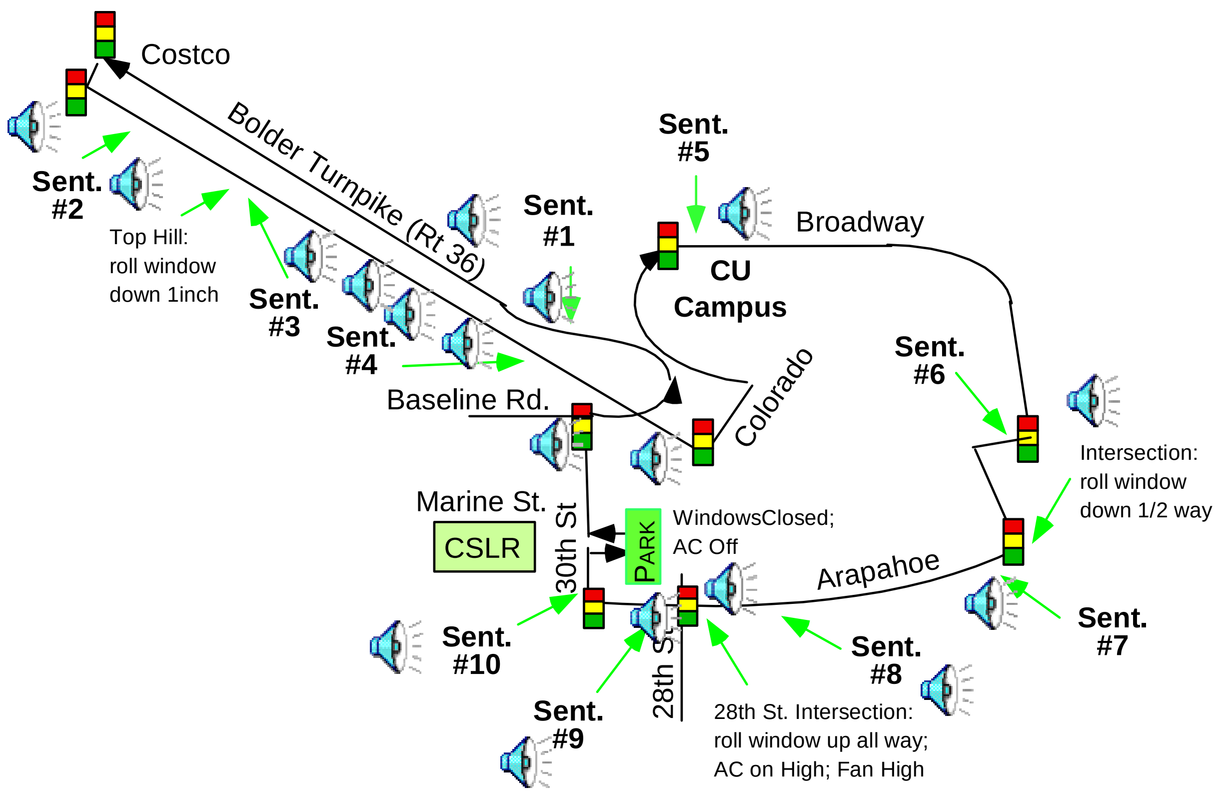


Figure 10: Boulder, CO route selected for Phase I data collection.

Below we summarize the 10 sentences produced (by J. Hansen, two examples of each) during Phase 1 collection. The text of these phonetically balanced sentences were obtained from the TIMIT speech corpus. These sentences were produced at the same location for each of the six GM vehicles during the approximately 17 mile route. This also helped significantly in tracking changes in route conditions for transcription (i.e., locations during the recording where turn signals or traffic noise conditions would be labeled for each vehicle).

```
1.  She had your dark suit in greasy wash water all year .
2.  Don't ask me to carry an oily rag like that
3.  Only the best players enjoy popularity.
4.  A good attitude is unbeatable.
5.  I honor my mom.
6.  Ambidextrous pickpockets accomplish more.
7.  Iguanas and alligators are tropical reptiles.
8.  Of course you can have another tunafish sandwich.
9.  An official deadline cannot be postponed.
10. Are you looking for employment?
```

Acoustic Noise Analysis

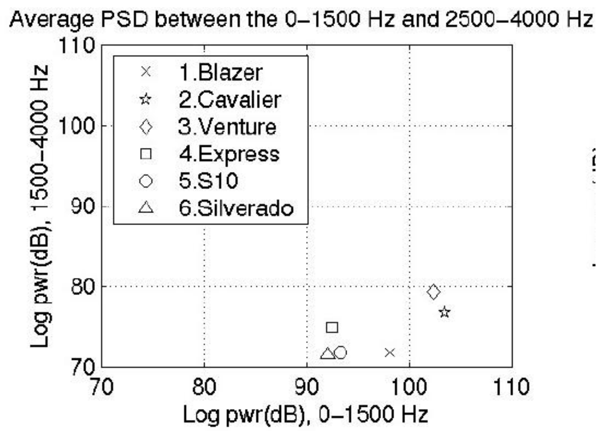
In this section, we summarize acoustic noise analysis of the CU-Move: Phase I data collection. All acoustic noise conditions are collected across 6 vehicles: Blazer, Cavalier, Venture, Express, S10, and Silverado. The noises were labeled into 14 categories which include:

1) Idle noise: the sound of the engine after starting and not moving, windows closed
2) Noise at 45 mph, window opened 1".
3) Noise at 45 mph, window closed.
4) Noise at 45 mph, window opened half way down.
5) Noise at 65 mph, window opened 1".
6) Noise at 65 mph, window closed.
7) Acceleration noise, window closed.
8) Acceleration noise, window opened half way down.
9) A/C (high) noise, window closed.
10) Deceleration noise, window opened 1".
11) Turn signal noise at 65 mph, window closed.
12) Turn signal noise, window opened 1".
13) Turn signal noise, window closed.
14) Wiper blade noise, window closed.

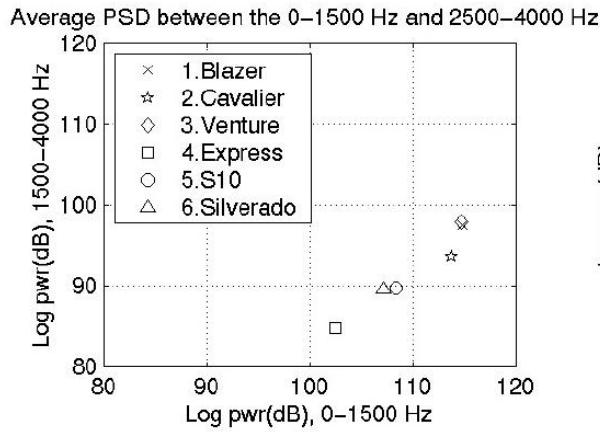
A total of 14 noise conditions were extracted from the same conditions and locations as possible for each of the 6 GM vehicles.

Figure 11 shows the scatter plots of 5 of the 14 noise conditions, including the average point, for all 6 cars. From these figures, we see that idle noise condition is located at the bottom point for all cars. The noise at 65 mph with window opened 1" occurs at the highest point for all cars. Most of noises show the same trend regarding low/high frequency location points for different cars. For example, the noise at 45 mph with windows closed and acceleration noise with windows closed are always close to the average PSD point for all cars. For the same car, the turn signal noise, windows opened 1" are always less than the windows closed condition, and both are less than the windows closed at 65 mph noise condition.

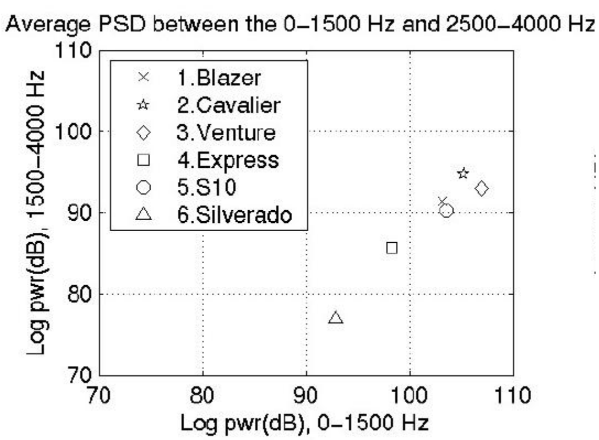
These scatter plots will help us to decide, in terms of the balance across noise PSD, which noises we can use to train for speech recognition in car environments in order to addressed all noise characteristics as possible. For example, car noise at 65 mph with windows opened 1" should be included as one necessary noise condition during speech recognition training, since it occupies an extreme point in the low/high frequency PSD space. In addition to selecting noise conditions that occupy extreme points in the low/high frequency PSD space, a representative sampling of noise conditions between these extremes should also be included. For the SUV (Blazer), one possible set of noise conditions to include would be (i) 65 mph with windows opened 1", (ii) noise at 45 mph with windows closed, (iii) turn signal noise with windows opened 1", and (iv) acceleration noise with windows opened half way.



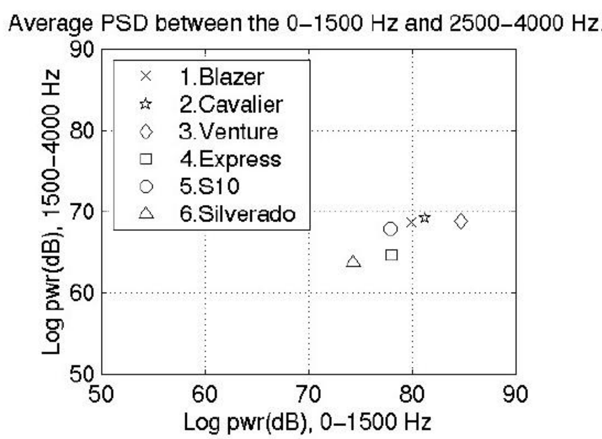
(a) Noise at 45 mph, window closed



(b) Noise at 65 mph, window opened 1"



(c) Acceleration noise, windows opened half way



(d) Idle noise, window closed

Figure 11: Summary of low versus high frequency log power in decibels for 4 noise conditions across 6 vehicles.

Other observations can be made based on the variability across vehicles. For example, if there is limited low/high frequency separation across the six vehicles, this would suggest that a common noise model could be constructed using noise from all vehicles. This would increase the available amount of training data. Some observations are as follows:

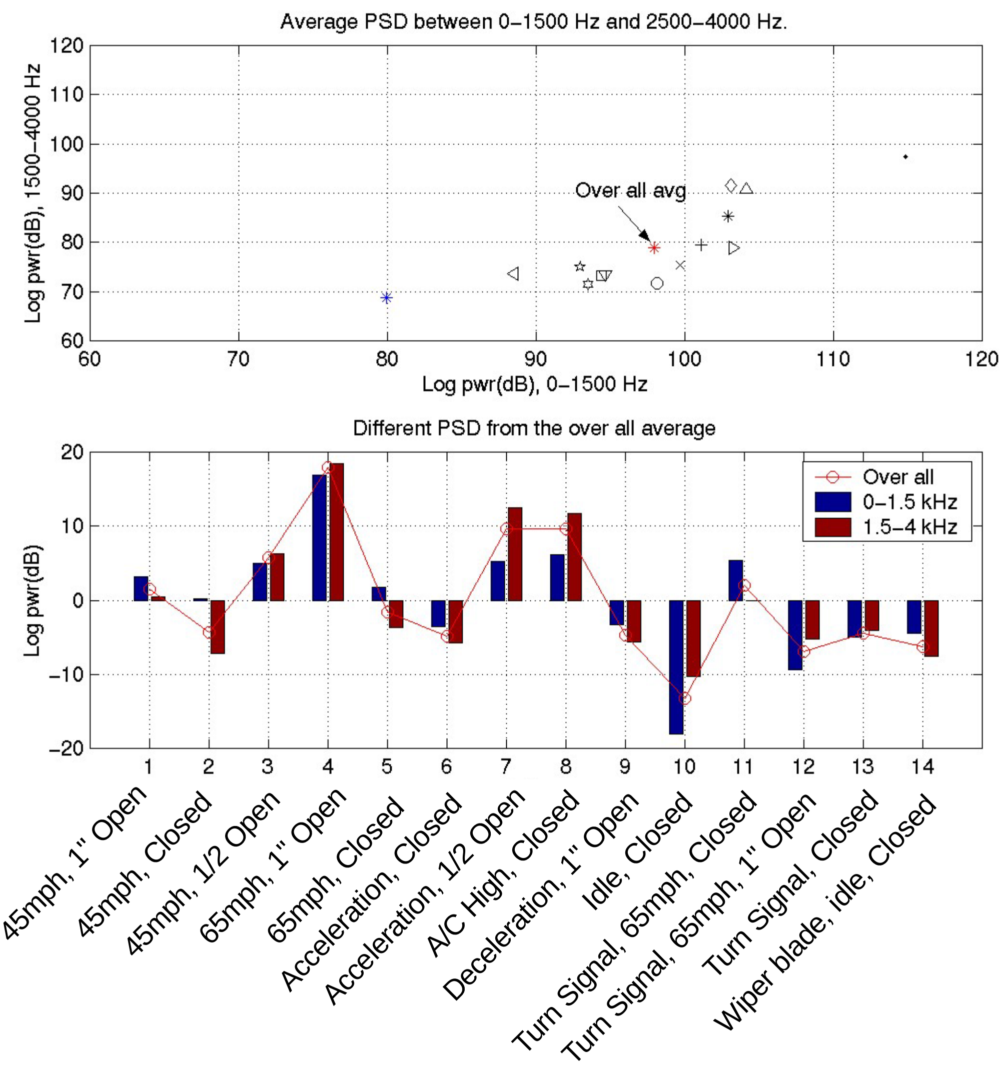
- **Idle Noise with Windows closed:** the six vehicles have similar noise levels. The EXP and SIL vehicles have slightly lower levels for the low frequency band, but the other four have about the same. Based on average noise energy levels, a single noise model would be sufficient for idle noise.
- **Wiper Blade Noise:** again we see that there is a slight separation across vehicles for noise energy in the low frequency band, but good agreement for high frequency band. A single noise model here with perhaps 2 mixtures would appear to work well.
- **Turn Signal Noise:** with windows closed, there appears to be a strong separation across the vehicles. This noise source would need to be uniquely modeled for each vehicle. These results vary when there is also **Windows Open Noise** included. For windows open 1 inch and turn signal, there was tight coupling of VEN, SUV, CAV and S10, EXP, SIL. Two different models would be needed here.
- **Deceleration with Windows Open 1 inch:** the CAV is the nosiest vehicle and the SIL is the most quiet. The other four vehicles grouped together.
- **Acceleration with Windows Open ½ way:** there is a wide noise level range for all six vehicles. Several models or multiple mixtures would be needed to cover the range of noise from the 6 vehicles. Similar observations can be made for acceleration with windows closed, though the spread across the vehicles is less.
- **65mph, Windows Closed:** there is good coupling for most of the vehicles. The CAV (compact car) and VEN (passenger van) were the most noisy.
- **65mph & 45mph, Windows open 1 inch:** both conditions showed a wider noise range for the six vehicles. Again, more than one noise model or mixture weight would be needed to represent this noise condition in speech recognition.
- **45mph, windows open ½ way:** this again showed a wide range of noise variation across the vehicles. The S10 was the most quiet, and the CAV was the most noisy.

GENERAL CONCLUSION: If we consider the relative balance between low (0-1500Hz) and high (1500-4000Hz) frequency energy content in PSD space for the six vehicles, we see that the CAV was generally the most noisy, the SUV was typically in the middle, and the SIL and S10 pickup trucks were the most quiet.

From the above scatter plots and other figures from our database, we can divide cars into 2 separate groups: SUV (Blazer) + CAV (Cavalier) + VEN (Venture) and EXP (Express) + S10 + SIL (Silverado). The first group was centered at higher energy content while the second group was concentrated at lower energy. We also find the SUV (Blazer) to be a lower bound for the upper group, and therefore located close to the average across the six vehicles. Therefore, the SUV (Blazer) would be a good selection since it is close to the mean of all 6 vehicles.

Symbol Notation for Figure 9 and numerical notation used for Figure 10.

1. Noise at 45 mph, window opened 1". (+)
2. Noise at 45 mph, window closed. (o)
3. Noise at 45 mph, window opened half way down. (* black)
4. Noise at 65 mph, window opened 1". (point)
5. Noise at 65 mph, window closed. (x)
6. Acceleration noise, window closed. (square)
7. Acceleration noise, window opened half way down. (diamond)
8. A/C (air conditioning, fan on high) noise, window closed. (upward pointing triangle)
9. Deceleration noise, window opened 1". (downward pointing triangle)
10. Idle noise: the sound of the engine after starting and no moving. (* blue)
11. Turn signal noise at 65 mph, window closed. (right pointing triangle)
12. Turn signal noise, window opened 1". (left pointing triangle)
13. Turn signal noise, window closed. (pentagram)
14. Wiper blade noise, window closed. (hexagram)



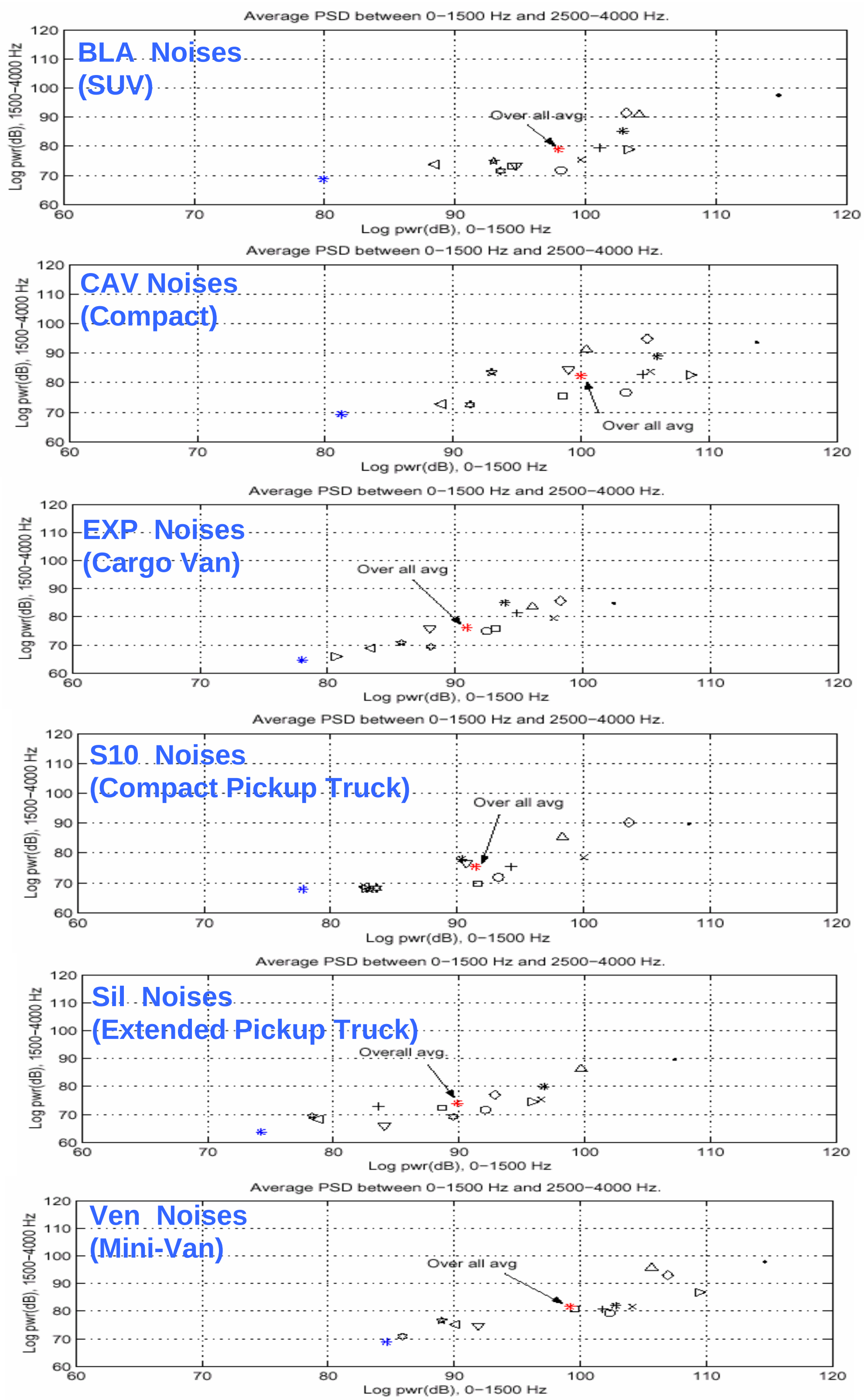


Figure 12: Scatter Plots of low vs. high Freq. 14 noise sources for each Vehicle

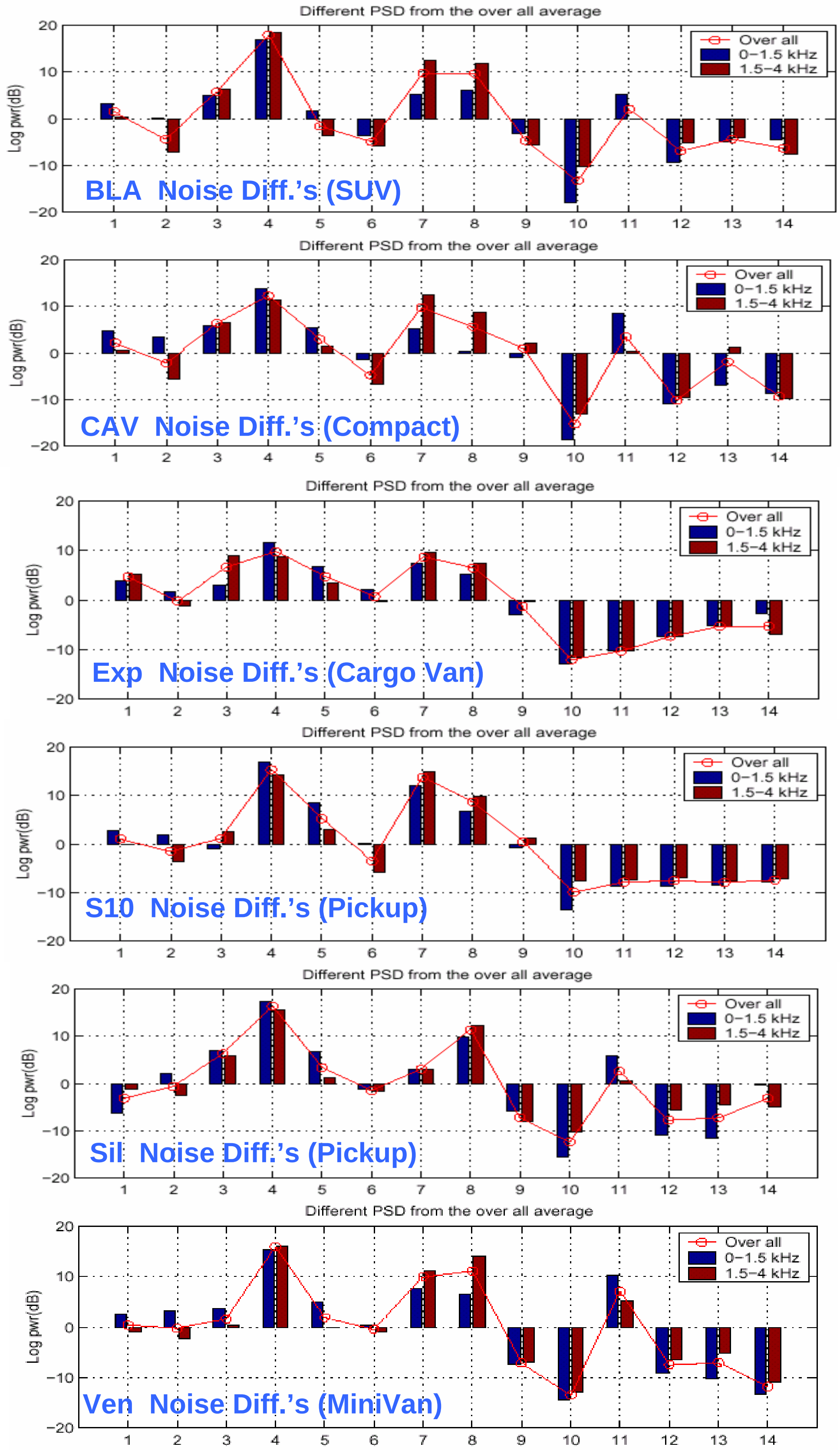


Figure 13: Deviation from Overall Avg. of 14 noise sources for each Vehicle

When we look at the scatterplots of low (0-1500Hz) versus high (1500-4000Hz) frequency content for the 14 noise conditions, the following observations are made. The noise condition "car idle" is consistently the lowest value across all six vehicles, with SIL pickup truck having the lowest noise value and VEN mini-van having the high noise value for idle. For the overall average noise value, (indicated with a red star) the VEN (mini-van) and CAV (compact) vehicles had the highest noise levels. The S10, SIL, and EXP all have similar overall average means, suggesting that a single noise model might be sufficient. The BLA (SUV) noise conditions were in general in the middle of the six vehicles (the VEN and CAV were generally noisier, and the S10, SIL, and EXP were generally slightly less noisy than the BLA). Figure 13 shows the deviation from overall average for each vehicle (note that the 14 noise conditions are numbered in order from 1. "idle noise" to 14. "wiper blade noise with windows closed". This figure allows us to directly compare noise level differences from overall mean to determine if one or more mixtures are needed for noise modeling in HMM speech recognition. We see similar trends for most of the noise conditions, with "11 .Turn signal noise at 65 mph, window closed" condition having more deviation from the vehicle means than other noises. We summarize specific recommendations on noise modeling for each noise conditions after a discussion on spectrogram analysis for a sample set of noises.

Spectrogram Analysis

For some noise conditions such as turn signal and wiper blade, it is more useful to consider time versus frequency plots to illustrate periodic characteristics. Figure 14 shows spectrograms of those noise conditions. The spectrograms clearly show the high frequency energy content of the turn signal noise for the vehicles (2-4kHz). The average impulse rate for the turn signal varied from 2.75 – 3.00Hz across the six vehicles, with sharp differences on how distinct the impulse points are from background noise.

<u>Turn Signal Analysis of Spectrograms:</u>

- **BLA**: SUV Blazer

<u>Turn signal with windows closed</u>: most of the energy is concentrated between 0 – 1.6 kHz, with periodic energy between 3.1– 4.0kHz at the following time locations: [0.1940, 0.5260, 0.8660, 1.200, 1.536, 1.87, 2.2060, 2.54, 2.8720] sec.

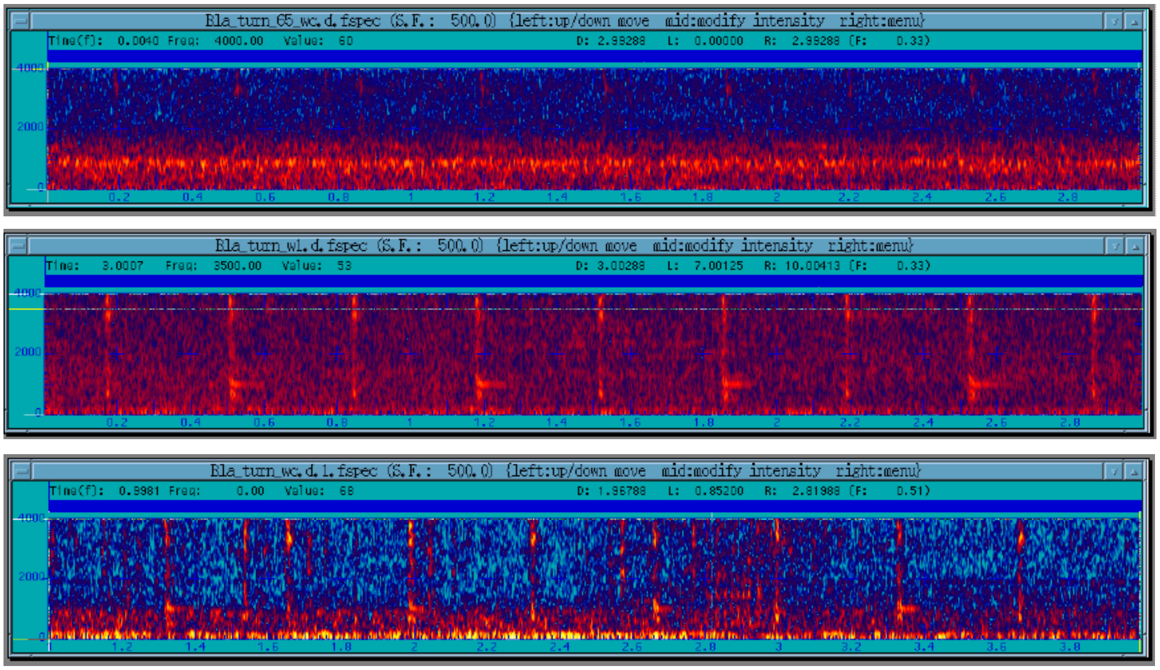
Avg. period = 0.2975 sec **F_0 = 3.361 Hz**

<u>Turn signal, 65mph with windows open 1-2inches:</u> periodic frequency concentration at the following time locations: [0.1720, 0.5140, 0.8500, 1.1880, 2.1940, 2.5320, 2.8680] sec.

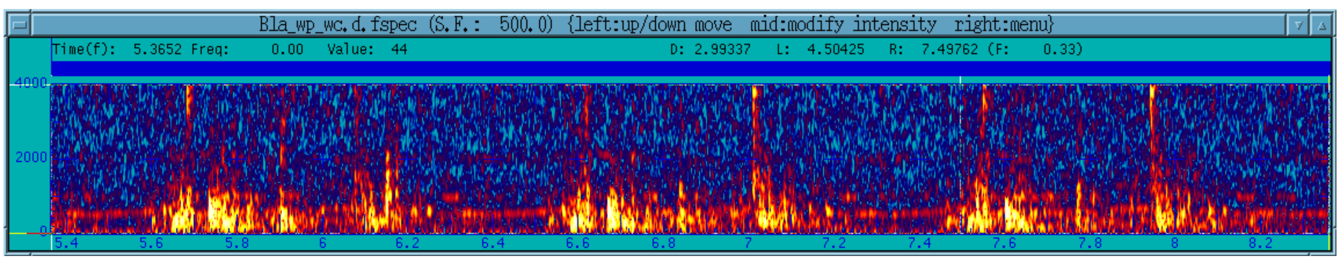
Avg. period = 0.3185 sec **F_0 = 3.140 Hz**

<u>Turn signal, 65mph with windows closed:</u> most of the energy is concentrated between 0:1000Hz, with periodic energy at the following time locations: [1.3221, 1.6581, 1.9921, 2.3281, 2.6641, 2.9961, 3.3321, 3.6681] sec.

Avg. period = 0.2932 sec **F_0 = 3.411 Hz**



BLA (SUV) Blazer: turn signal with (i) windows closed, 65mph, (ii) windows open 1-2 inches, (iii) windows closed



Blazer, wiper blade, window closed.

Figure 14: Spectrographic analysis of wiper blade noise across SUV vehicle

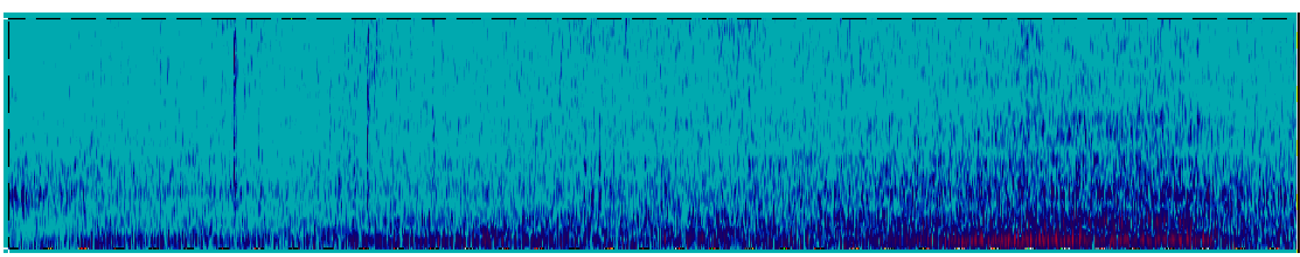


Figure 15: Spectrographic analysis of BLA for acceleration onto a highway.

The results from this section clearly show a range of noise conditions for the six vehicles selected. While there is variability across the vehicles, the Chevy Blazer SUV represents one vehicle in the middle of noise distribution scatter plots. The CAV and VEN were on average noisy than SUV, while the EXP, S10, SIL were slightly more quite. As a result, we selected SUV as the vehicle for Phase II data collection across the United States.

Power Spectrum Density (PSD) with Low vs. High Frequency Band Comparisons

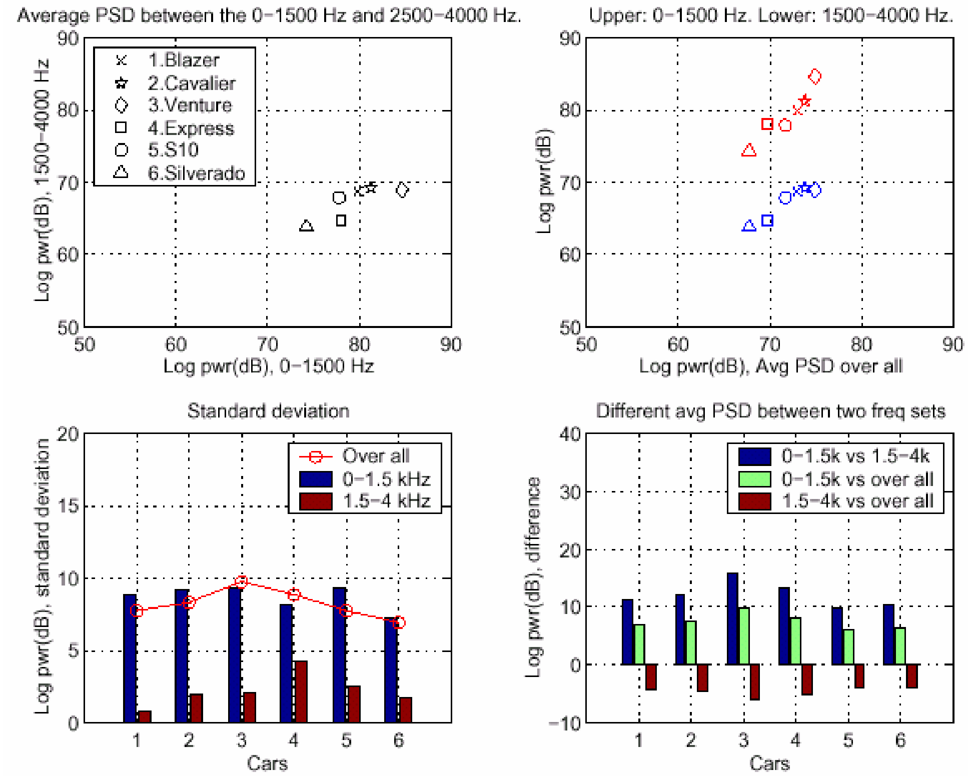


Figure 16: Engine Idle noise (car not moving), windows closed

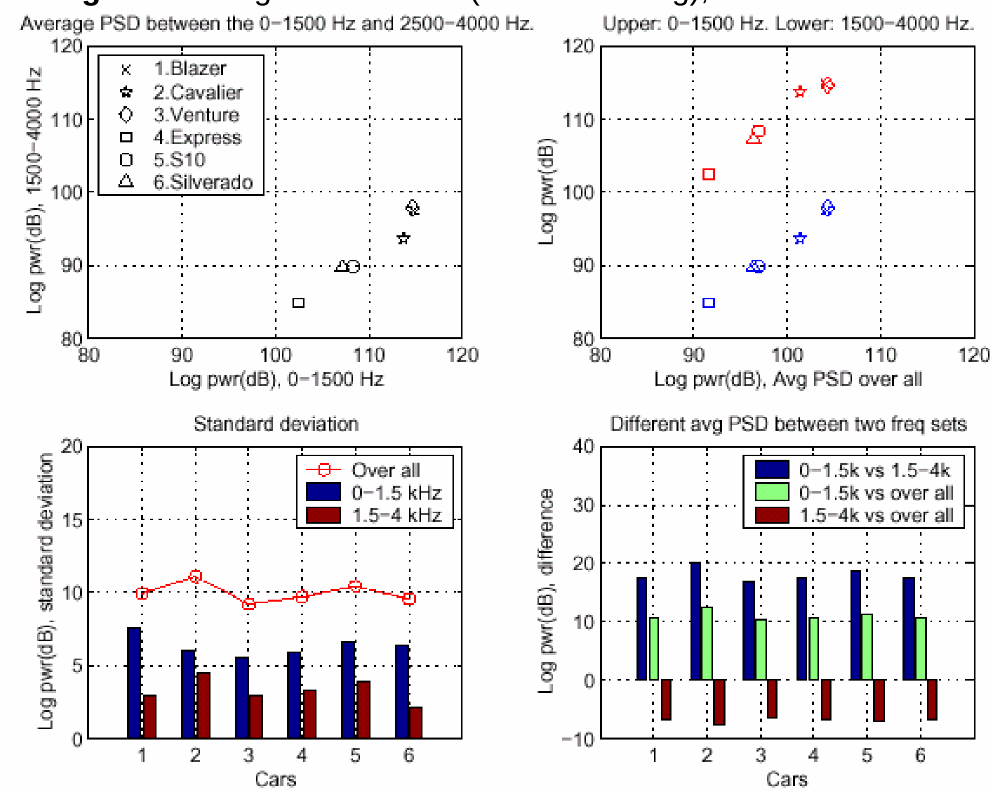


Figure 17: Noise at 65 mph, window open 1 inch

Section 5: Speaker Information

In this section, we focus on summarizing details on the speaker population that make up the CU-Move Corpus. Recall that speech was collected from six U.S. cities as shown in Fig. 18 below.

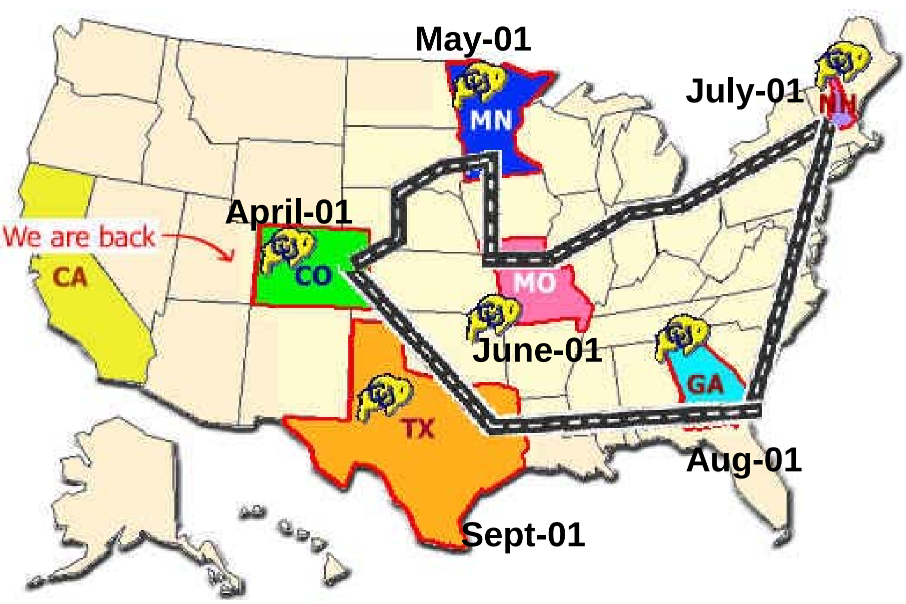


Figure 18: United States Map of cities visited and dates where speech was collected.

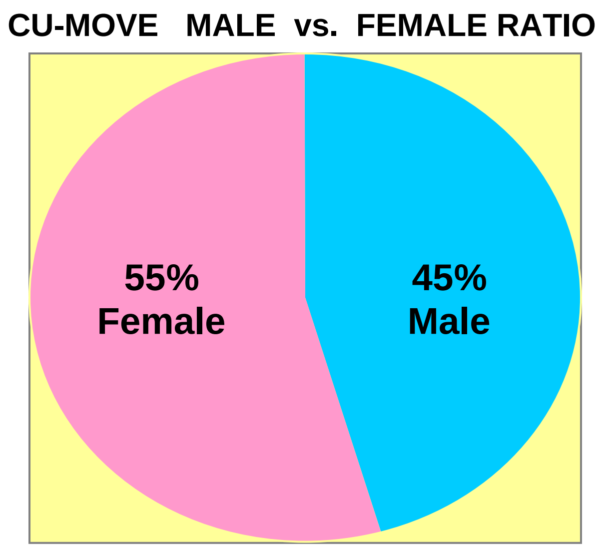
The complete CU-Move corpus consists of 3 hard disk distributions (60GB,80GB,40GB). The 180GB of data represents a downsampled, beamformed, and organized version from our original +700GB of data originally sampled at 44.1kHz. The following lists which speakers and cities are associated with which disks.

Disk #1:	Minneapolis, MN:	Speakers 0050 -- 0205
Disk #2:	St. Louis, MO:	Speakers 0206 -- 0360
	Manchester, NH:	Speakers 0361 -- 0485
Disk #3:	Savannah, GA:	Speakers 0486 -- 0517
	Dallas, TX:	Speakers 0519 -- 0548
	Boulder, CO:	Speakers 0001 -- 0028

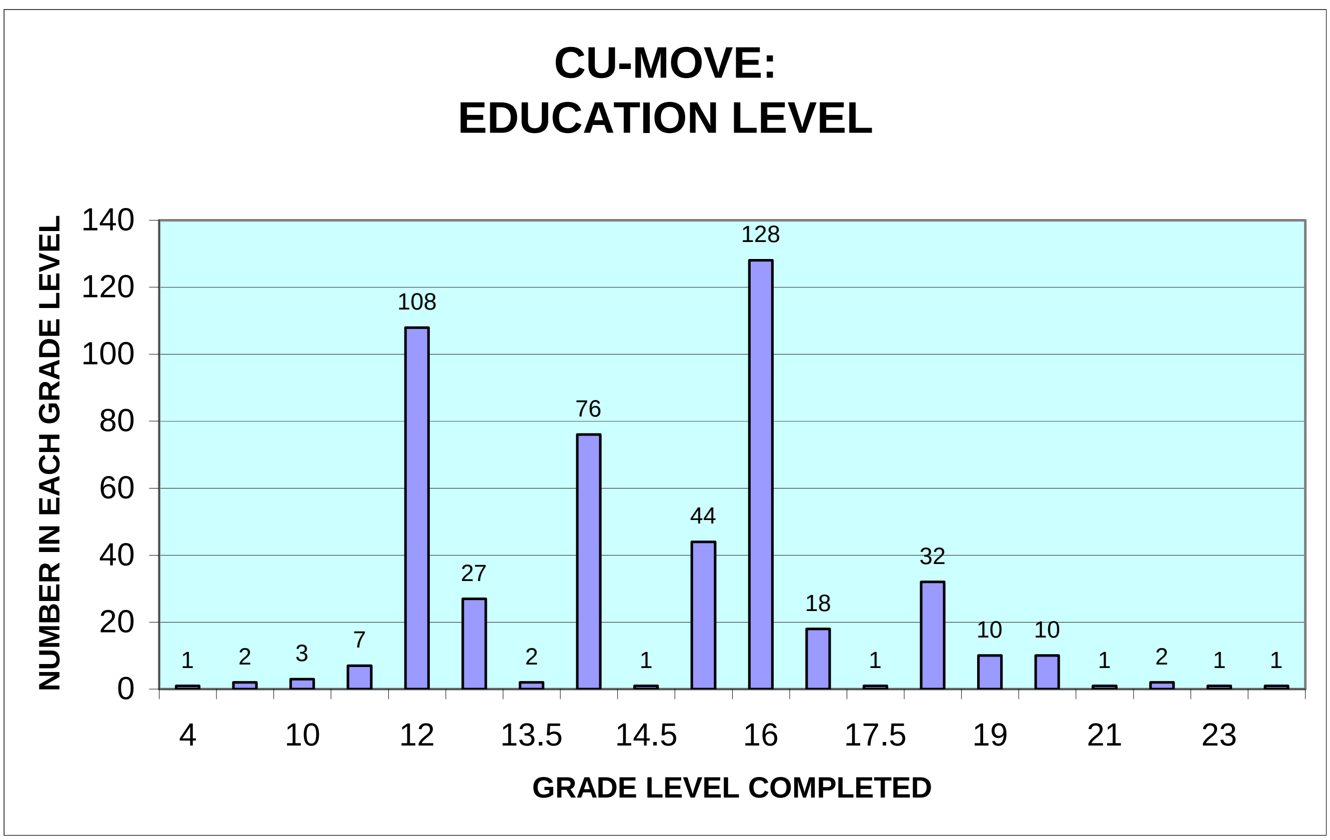
In this section, we summarize the per city speaker information listings, as well as provide the recommended (i) train, (ii) development test, and (iii) test sets for the CU-Move corpus.

Speaker Information on CU-Move:

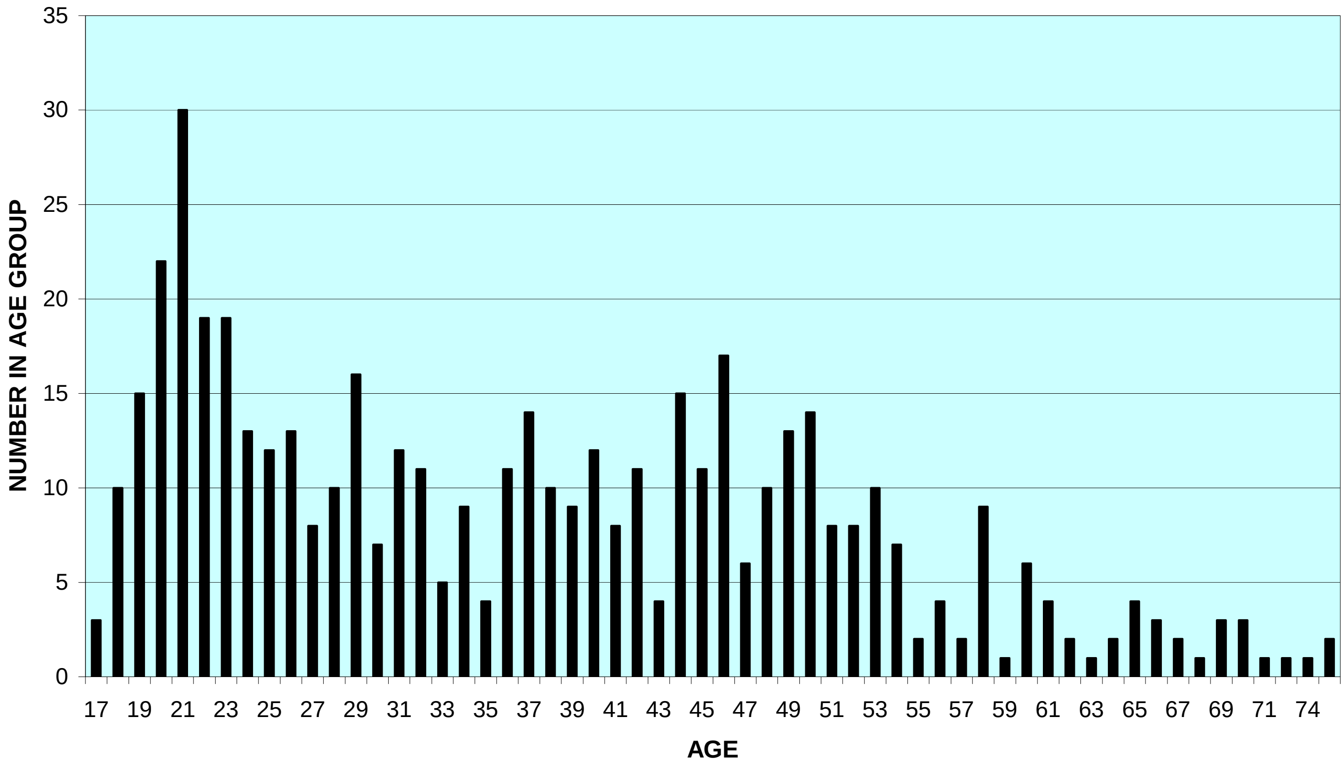
- **Male/Female Distribution:**



- **Educational Level (grade level completed):**

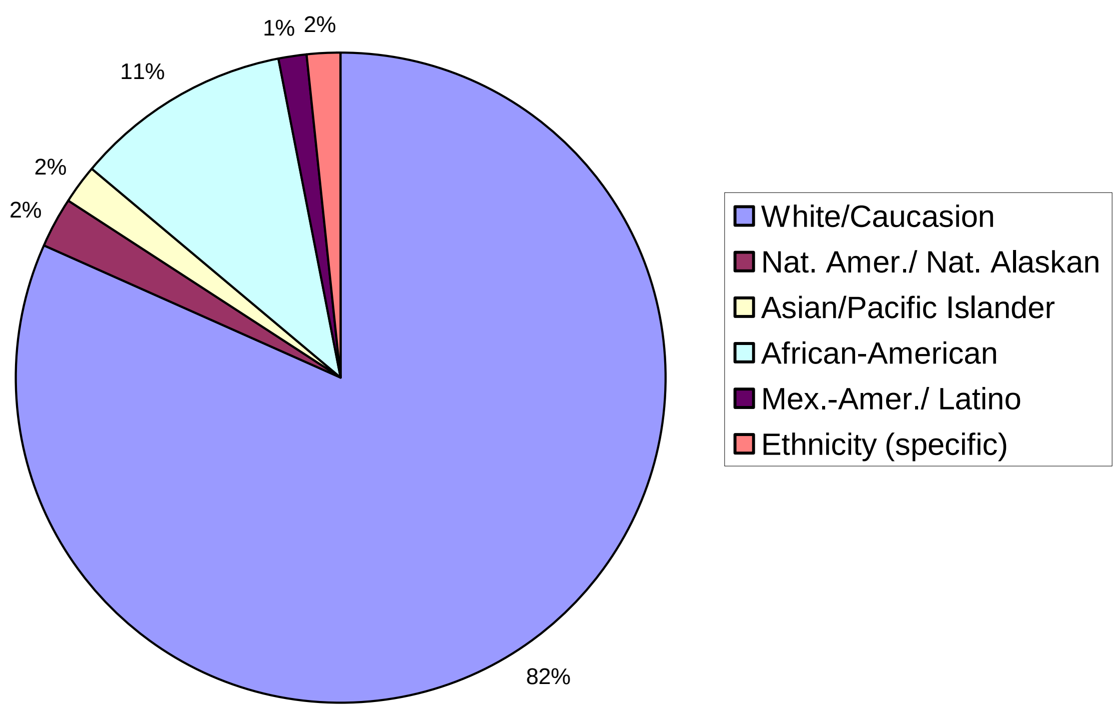


- **<u>Age Level Distribution CU-Move Corpus:</u>**



- **<u>CU-Move: Ethnicity of Participants:</u>**

White / Caucasion	Nat. Amer./ Nat. Alaskan	Asian/Pacific Islander	African American	Mex.-Amer./ Latino	Ethnicity (specific)
402	12	9	54	7	8



- ## Speaker Lists: Train, Development-Test, Test sets

Here, we summarize the recommended training, development-test, and test speaker sets for the CU-Move corpus across the six cities. The lists are provided in terms of increasing age (from 17-75 years old), and separated between male and female lists. The format is as follows:

Format for lists: Speaker Number Age{yrs}

So, for Boulder, CO, the first male speaker in the training set is speaker 15, who is 20 years old. The first female is speaker 16, who is 22 years old.

Boulder, CO {Speakers 1-28}

MALE				Development		FEMALE		Development			
Training Set		Test Set		Test Set		Training Set		Test Set		Test Set	
(6 spkrs)		(2 spkrs)		(5 spkrs)		(7 spkrs)		(2 spkrs)		(6 spkrs)	
15	20	19	21	[28]	21	16	22	[7]	24	[5[	27
[3]	24			[10]	24	12	27			20	29
[6]	25	[1]	28	[2]	29	22	29	[4]	31	18	31
[8]	29					14	41			17	46
[25]	45			23	46	[9]	49			21	50
[27]	49			13	70	24	61			11	66
						[26]	72				

Boulder, CO Note: the speakers denoted [x] represent missing speakers [1-10, 25-28]. A number of initial speakers from Boulder volunteered during an initial training and testing phase for the team going on the road. At this time, these speakers are not available. Any changes in the status of these speakers will be posted on the CU-Move web page.

Savannah, GA {Speakers 486-517}

MALE		Development				FEMALE		Development			
Training Set		Test Set		Test Set		Training Set		Test Set		Test Set	
(8 spkrs)		(3 spkrs)		(6 spkrs)		(7 spkrs)		(3 spkrs)		(5 spkrs)	
495	19	494	20	515	20	502	19	505	21		
516	21					490	22			496	22
493	25	503	25	507	25	491	26	497	29		
492	29					506	29			517	31
488	35	510	35	511	35	504	34			489	38
487	36			501	42	486	43	512	45	509	46
498	43			508	43	500	53			499	69
513	43			514	48						

Savannah, GA Note: all speakers present, verified, and okay.

Dallas, TX {Speakers 519-548}

MALE		Development				FEMALE		Development			
Training Set		Test Set		Test Set		Training Set		Test Set		Test Set	
(7 spkrs)		(2 spkrs)		(5 spkrs)		(8 spkrs)		(3 spkrs)		(5 spkrs)	
538	18			520	22	532	21			523	22
548	25			533	26	522	32	527	32	528	33
519	32	524	37	536	37	541	36	530	37	540	38
547	43			534	53	526	39				
543	58			535	60	537	39			544	40
[525]	62	531	65			545	45	542	48		
546	66					521	54			529	58
						539	64				

Dallas, TX Note: Speaker 525 is incomplete, but all other speakers present, verified, and okay.

Manchester, NH {Speakers 361-485}

MALE						**FEMALE**					
Training Set		**Development Test Set**		**Test Set**		**Training Set**		**Development Test Set**		**Test Set**	
(26 spkrs)		(9 spkrs)		(18 spkrs)		(35 spkrs)		(12 spkrs)		(25 spkrs)	
402	18	408	18	366	19	403	18	426	18	448	18
369	19					449	18				
380	20	383	20	391	20	433	19				
392	20	362	21	395	22	411	20	416	20	430	20
387	23			400	23	441	20				
452	26			454	26	444	22	393	23	427	23
446	27					365	24	412	24	415	24
390	28					459	24			405	25
425	28	462	28	396	29	442	25				
480	29					378	26			472	26
361	30	[414]	30	456	30	388	27			410	29
406	31			[420]	31	434	29	413	31	419	31
481	31					468	31				
418	32	469	32	482	32	367	32			428	32
368	35	363	36	485	36	404	33	[435]	33		
423	37					451	37			461	38
379	38			397	38	398	40	421	40	464	40
447	39					477	40				
478	40			483	40	375	41			465	41
374	42			484	42	381	42	399	42	445	42
457	44	382	45	471	45	424	43				
474	45					376	44			409	44
437	46			463	48	443	44	458	44	460	44
371	49	453	50	455	50	470	44				
422	51			450	52	439	45			364	46
384	53			394	69	466	46			479	46
						429	47	389	48	436	48
						386	49			475	49
						370	50	373	50	432	50
						438	50				
						407	51			467	51
						473	51			417	52
						385	53			377	54
						401	54			476	54
						372	58	431	60	[440]	65

Manchester, NH Note: the speakers denoted [x] represent missing speakers [414,420,435,440]. These speakers were either (i) originally collected, but were incomplete, or (ii) there were troubles in complete transfer of data from our digital recorder hard-disk system unit to PC. Any changes in the status of these speakers will be posted on the CU-Move web page.

Minneapolis, MN {Speakers 50-205}

MALE						**FEMALE**					
Training Set		**Development Test Set**		**Test Set**		**Training Set**		**Development Test Set**		**Test Set**	
(39 spkrs)		(12 spkrs)		(23 spkrs)		(43 spkrs)		(12 spkrs)		(26 spkrs)	
185	17	66	18	183	19	63	17	196	19	202	19
189	19					132	20			167	21
198	19			134	20	201	21	203	21		
137	20					130	22				
153	20	171	20			142	22	165	22	176	22
131	21			174	21	182	22			195	22
205	21			192	22	166	23	188	23	204	24
91	23	107	23	[133]	23	80	26				
175	23					173	28	143	29		
184	23			199	23	172	29			139	31
[135]	24			154	24	194	33				
181	25					56	34				
89	26			152	26	[129]	34	159	34	164	34
149	28			197	28	169	34				
151	29	124	30	136	30	95	36			100	36
191	32					60	37			186	37
187	34	122	35	62	36	68	39	112	40		
110	37					123	40			200	40
157	37			71	38	61	41				
155	38	158	38	87	39	150	42			72	44
90	40					75	44				
114	40	160	41	190	41	96	44	106	44	115	44
88	42			103	42	118	44				
145	42					161	45				
101	44	121	45	141	45	57	46	58	46	98	46
128	46					126	46				
178	48	59	49	92	50	146	46			168	46
94	50					67	47				
108	50					109	47			144	47
102	52	119	52	125	52	50	48			77	48
82	53					127	48				
97	53			138	57	78	49			83	49
117	58					170	49				
65	59	84	60	156	61	140	50			180	50
74	62					147	51	162	51	179	51
120	63	52	65	93	66	79	52				
53	69					81	53			116	53
55	70			104	71	76	54			99	54
69	77			113	77	148	56			193	56
						86	58			73	60
						163	60	51	61	70	64
						[85]	67	105	67	177	69
						54	70			139	xx

Minneapolis, MN Note: the speakers denoted [x] represent missing speakers [85,129,133,135]. These speakers were either (i) originally collected, but were incomplete, or (ii) there were troubles in complete transfer of data from our digital recorder hard-disk system unit to PC. Any changes in the status of these speakers will be posted on the CU-Move web page. Speakers {64,111} did not have speaker information/release forms available and therefore are not included in documentation or listings. Finally, female Speaker #139 did not provide her birthdate or age, so she is listed as "xx" under age.

St. Louis, MO {Speakers 206-360}

MALE Training Set (40 spkrs)		Development Test Set (10 spkrs)		Test Set (19 spkrs)		FEMALE Training Set (45 spkrs)		Development Test Set (12 spkrs)		Test Set (28 spkrs)	
207	17	273	18	274	19	268	18	216	19	224	19
323	19					[298]	19				
209	20	284	20	[288]	20	317	19			320	19
343	20					208	20	[230]	20	248	20
252	21			263	21	255	20				
282	21					212	21	218	21	221	21
[294]	21					222	21			223	21
311	21	318	21	322	21	226	21			234	21
251	22					257	21				
316	22			319	22	258	21				
337	22					261	21	[293]	21		
275	23	[286]	23	[291]	23	312	21			327	21
332	23					228	22	237	22	300	22
307	24			309	24	243	23			279	23
220	25					[302]	23	[295]	24	227	25
338	25			[290]	27	242	25				
331	27					246	25			[299]	25
313	28			330	28	[303]	25				
232	29	238	29	254	29	219	26	231	26	253	26
342	29			349	31	278	26			334	26
215	32					270	27			[297]	27
232	34			328	35	249	28			213	29
329	35	[289]	36	314	37	272	30				
315	37					280	30			256	31
358	38			247	39	[321]	31				
[287]	41	348	41			341	32			281	33
217	43			305	43	[296]	34				
245	44					310	34	211	36	250	36
229	46	240	47			276	36				
355	47					277	36			333	37
235	49			336	49	346	37				
350	50					356	37	271	38	[292]	38
283	52	326	52			244	39			[301]	39
360	52			211	53	335	39			265	42
269	53					308	42				
339	55			345	55	236	45	266	45	225	46
267	57					262	46				
351	58			352	58	325	46				
324	60	259	61			[304]	48	344	48	264	49
354	68					[285]	49				
						210	50			306	51
						214	53	206	54	260	56
						347	56			340	58
						359	58			357	65
						353	74				

St. Louis, MO Note: the speakers denoted [x] represent missing speakers [230,239,285-304,321]. These speakers were either (i) originally collected, but were incomplete, or (ii) there were troubles in complete transfer of data from our digital recorder hard-disk system unit to PC (this occurred for speaker set 285-304). Any changes in the status of these speakers will be posted on the CU-Move web page.

CU-Move
Speaker Information

Minniapolis, MN

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/Caucasion	Native American/ Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/Latino	Ethnicity (specific)
50	48	102952	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
51	61	71539	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
52	65	110435	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
53	69	122031	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
54	70	30531	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
55	70	101330	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
56	34	32767	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
57	46	120854	FALSE	TRUE	16	TRUE	TRUE	FALSE	FALSE	FALSE	
58	46	82754	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
59	49	12952	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
60	37	70563	FALSE	TRUE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
61	41	13160	FALSE	TRUE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
62	36	51865	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
63	17	72783	FALSE	TRUE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
65	59	100741	TRUE	FALSE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
66	18	42983	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
67	47	51354	FALSE	TRUE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
68	39	110017	FALSE	TRUE	16	FALSE	FALSE	TRUE	FALSE	FALSE	
69	77	20724	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
70	64	70536	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
71	38	30663	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
72	44	80856	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
73	60	120640	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
74	62	82438	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
75	44	82056	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
76	54	52747	FALSE	TRUE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
77	48	43054	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
78	49	101051	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
79	52	122148	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
80	26	112374	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
81	53	121147	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
82	53	100447	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
83	49	111651	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
84	60		TRUE	FALSE		FALSE	FALSE	FALSE	FALSE	FALSE	
85	67	100733	FALSE	TRUE	16	FALSE	FALSE	TRUE	FALSE	FALSE	
86	58	122942	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
87	39	30662	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
88	42	92558	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
89	26	11775	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
90	40	80160	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	

91	23	71227	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
92	50	32849	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
93	66	11735	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
94	50	10151	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE
95	36	22365	FALSE	TRUE	16	FALSE	TRUE	FALSE	FALSE	FALSE
96	44	72656	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE
97	53	82947	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE
98	46	92354	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE
99	54	52247	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE
100	36	61165	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE
101	44	111056	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
102	52	122748	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
103	42	53059	TRUE	FALSE	16	FALSE	FALSE	FALSE	FALSE	TRUE
104	71	82529	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE
105	67	71833	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE
106	44	22557	FALSE	TRUE	22	TRUE	FALSE	FALSE	FALSE	FALSE
107	23	102577	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
108	50	40951	TRUE	FALSE	18	FALSE	FALSE	FALSE	TRUE	FALSE
109	47	113053	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE
110	37	81164	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
112	40	41861	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE
113	77	102823	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE
114	40	102160	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
115	44	120756	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE
116	53	32148	FALSE	TRUE	19	TRUE	FALSE	FALSE	FALSE	FALSE
117	58	121442	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE
118	44	33056	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE
119	52	72848	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
120	63	120301	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE
121	45	72255	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
122	35	30966	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE
123	40	80560	FALSE	TRUE	16	FALSE	FALSE	FALSE	FALSE	FALSE
124	30	20671	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE
125	52	60149	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE
126	46	90454	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE
127	48	62753	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE
128	46	90154	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
129	34	82066	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE
130	22	92278	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE
131	21	12180	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE
132	20	61781	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE
133	23	30778	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
134	20	101580	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE
135	24	0	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE
136	30	20371	TRUE	FALSE	13	FALSE	FALSE	FALSE	TRUE	FALSE
137	20	101580	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE
138	57	111943	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE

139	31	32470	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
139			FALSE	TRUE		FALSE	FALSE	FALSE	FALSE	FALSE	
140	50	41951	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
141	45	40456	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
142	22	30879	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
143	29	101271	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
144	47	0	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
145	42	10759	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
146	46	80555	FALSE	TRUE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
147	51	92149	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
148	56	103044	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
149	28	120672	TRUE	FALSE		FALSE	FALSE	FALSE	TRUE	FALSE	
150	42	61459	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
151	29	12572	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
152	26	61175	TRUE	FALSE	12	FALSE	FALSE	FALSE	FALSE	FALSE	HYBRID
153	20	100989	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
154	24	31977	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
155	38	41762	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
156	61	92534	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
157	37	32464	TRUE	FALSE	9	TRUE	FALSE	FALSE	FALSE	FALSE	
158	38	60263	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
159	34	122166	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
160	41	52360	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
161	45	101955	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
162	51	71949	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
163	60	30741	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
164	34	71766	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
165	22	110578	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
166	23	102077	FALSE	TRUE	16	FALSE	FALSE	FALSE	FALSE	FALSE	BLACK & WHITE
167	21	122179	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
168	46	30755	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
169	34	32567	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
170	49	112251	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
171	20	80380	TRUE	FALSE	15	FALSE	FALSE	TRUE	FALSE	FALSE	
172	29	41772	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
173	28	11873	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
174	21	22980	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
175	23	91477	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
176	22	50279	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
177	69	71231	FALSE	TRUE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
178	48	111752	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
179	51	81949	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
180	50	121050	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	

181	25	120275	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
182	22	100778	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
183	19	11882	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
184	23	62378	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
185	17	61184	TRUE	FALSE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
186	37	32064	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
187	34	120966	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
188	23	90677	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
189	19	100581	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
190	41	91859	TRUE	FALSE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
191	32	40869	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
192	22	90378	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
193	56	111544	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
194	333	30368	FALSE	TRUE	16	FALSE	TRUE	FALSE	FALSE	FALSE	
195	22	123078	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
196	19	20282	FALSE	TRUE	13	FALSE	FALSE	FALSE	FALSE	FALSE	ASIAN/WHITE
197	28	110172	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
198	19	91981	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
199	23	52078	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
200	40	120660	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
201	21	31880	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
202	19	81381	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
203	21	62701	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
204	24	90576	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
205	21	30780	TRUE	FALSE	15	FALSE	FALSE	FALSE	TRUE	FALSE	

St. Louis, MO

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/Caucasion	Native American/Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/Latino	Ethnicity (specific)
206	54	30247	FALSE	TRUE	16	FALSE	TRUE	FALSE	FALSE	FALSE	
207	17	41884	TRUE	FALSE	10	TRUE	FALSE	FALSE	FALSE	FALSE	
208	20	11181	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
209	20	32381	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
210	50	122350	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
211	53	12148	TRUE	FALSE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
212	21	70480	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
213	29	81271	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
214	53	121347	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
215	32	80768	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
216	19	101281	FALSE	TRUE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
217	43	41258	TRUE	FALSE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
218	21	81079	FALSE	TRUE	11	FALSE	FALSE	FALSE	TRUE	FALSE	
219	26	22575	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
220	25	83175	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
221	21	30780	FALSE	TRUE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
222	21	21980	FALSE	TRUE	15	FALSE	FALSE	TRUE	FALSE	FALSE	
223	21	40880	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
224	19	120481	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
225	46	10155	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
226	21	50280	FALSE	TRUE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
227	25	20976	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
228	22	120378	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
229	46	80454	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
230	20	11881	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
231	26	10975	FALSE	TRUE	18	FALSE	FALSE	FALSE	TRUE	FALSE	
232	29	71301	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
232	34	102166	TRUE	FALSE	19	FALSE	FALSE	FALSE	FALSE	FALSE	HUMAN
234	21	71479	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
235	49	71551	TRUE	FALSE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
236	45	30756	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
237	22	111978	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
238	29	32472	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
240	47	110953	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
211	36	10865	FALSE	TRUE	22	TRUE	FALSE	FALSE	FALSE	FALSE	
242	25	90375	FALSE	TRUE	16	FALSE	FALSE	TRUE	FALSE	FALSE	
243	23	12578	FALSE	TRUE	17	FALSE	FALSE	TRUE	TRUE	FALSE	
244	39	82961	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
245	44	70257	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
246	25	30476	FALSE	TRUE	16	TRUE	TRUE	FALSE	FALSE	FALSE	
247	39	12562	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
248	20	60281	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
249	28	91272	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
250	36	122864	FALSE	TRUE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
251	22	80278	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
252	21	51280	TRUE	FALSE	15	FALSE	FALSE	FALSE	TRUE	FALSE	

253	26	42075	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
254	29	111371	TRUE	FALSE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
255	20	51981	FALSE	TRUE	15	FALSE	FALSE	FALSE	FALSE	TRUE	
256	31	81069	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
257	21	41480	FALSE	TRUE	12	FALSE	FALSE	FALSE	FALSE	TRUE	
258	21	120679	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
259	61	52040	TRUE	FALSE	14	FALSE	FALSE	FALSE	FALSE	FALSE	OTHER
260	56	70545	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
261	21	50380	FALSE	TRUE	15	FALSE	FALSE	FALSE	FALSE	FALSE	OTHER
262	46	82554	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
263	21	72879	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
264	49	52452	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
265	42	40559	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
266	45	92755	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
267	57	101243	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
268	18	91882	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
269	53	91842	TRUE	FALSE	23	TRUE	FALSE	FALSE	FALSE	FALSE	
270	27	21474	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
271	38	501634	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
272	30	121570	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
273	18	112382	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
274	19	92281	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
275	23	70528	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
276	36	82464	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
277	36	62165	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
278	26	40975	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
279	23	60878	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
280	30	50171	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
281	33	30468	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
282	21	12180	TRUE	FALSE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
283	52	10445	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
284	20	12981	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
285	49	60352	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
286	23	71901	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
287	41	92059	TRUE	FALSE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
288	20	21081	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
289	36	21065	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
290	27	22574	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
291	23	81777	TRUE	FALSE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
292	38	72362	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
293	21	43080	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
295	24	71977	FALSE	TRUE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
294	21	62480	TRUE	FALSE		TRUE	FALSE	FALSE	FALSE	FALSE	
296	34	61667	FALSE	TRUE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
297	27		FALSE	TRUE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
298	19	111081	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
299	25	71901	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
300	22	123078	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
301	39	110861	FALSE	TRUE	21	FALSE	FALSE	FALSE	FALSE	FALSE	JEWISH
302	23	72178	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
303	25	32476	FALSE	TRUE	17	FALSE	FALSE	TRUE	FALSE	FALSE	
304	48	40653	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	

305	43		TRUE	FALSE		TRUE	FALSE	FALSE	FALSE	FALSE	
306	51	80749	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
307	24	83176	TRUE	FALSE	0	TRUE	FALSE	FALSE	TRUE	FALSE	
308	42	31359	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
309	24	62177	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
310	34	11067	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
311	21	92179	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
312	21	11580	FALSE	TRUE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
313	28	80672	TRUE	FALSE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
314	37	11364	TRUE	FALSE	17	FALSE	FALSE	FALSE	TRUE	FALSE	
315	37	51564	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
316	22	110578	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
317	19	40282	FALSE	TRUE	14	FALSE	FALSE	FALSE	FALSE	TRUE	
318	21	50280	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
319	22	20479	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
320	19	102681	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
321	31	81869	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
322	21	40480	TRUE	FALSE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
323	19	90181	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
324	60	61241	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
325	46	112954	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
326	52	83148	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
327	21	120879	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
328	35	91565	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
329	35	71166	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
330	28	21273	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
331	27	22674	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
332	23	32478	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
333	37	92063	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
334	26	11873	FALSE	FALSE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
335	39	102361	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
336	49	61452	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
337	2	12178	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
338	25	61376	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
339	55	12546	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
340	58	82447	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
341	32	31869	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
342	29	112371	TRUE	FALSE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
343	20	110580	TRUE	FALSE	15	FALSE	FALSE	FALSE	FALSE	FALSE	ARAB
344	48	11053	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
345	55	112645	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
346	37	21164	FALSE	TRUE	25	TRUE	FALSE	FALSE	FALSE	FALSE	
347	56	52145	FALSE	FALSE	0	FALSE	FALSE	FALSE	FALSE	FALSE	
348	41	20160	TRUE	FALSE	16	FALSE	TRUE	FALSE	FALSE	FALSE	
349	31	61170	TRUE	FALSE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
350	50	72151	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
351	58	21643	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
352	58	71201	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
353	74	11927	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
354	68	30932	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
355	47	80701	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
356	37	40364	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	

357	65	70836	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
358	38	61163	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
359	58	20643	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
360	52	91548	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE	

Manchester, NH

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/ Caucasion	Native American/Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/ LatiFALSE	Ethnicity (specific)
361	30	92876	TRUE	FALSE	15	FALSE	TRUE	FALSE	FALSE	FALSE	
362	21	101179	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
363	36	110564	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
364	46	62855	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
365	24	71877	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
366	19	21782	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
367	32	20769	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
368	35	100665	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
369	19	11	TRUE	FALSE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
370	50	12051	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
371	49	41052	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
372	58	90342	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
373	50	102350	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
374	42	11059	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
375	41	62060	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
376	44	80757	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
377	54	81101	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
378	26	41175	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
379	38	10763	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
380	20	72081	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
381	42	111058	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
382	45	81555	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
383	20	82480	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
384	53	102447	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
385	53	33048	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
386	49	80652	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
387	23	122977	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
388	27	102473	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
389	48	111052	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
390	28	101172	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
391	20	112980	TRUE	FALSE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
392	20	71181	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
393	23	51578	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
394	69	61332	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
395	22	61579	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	

396	29	110571	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
397	38	90362	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
398	40	80561	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
399	42	111358	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
400	23	100978	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
401	54	92546	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
402	18	51683	TRUE	FALSE	11	TRUE	FALSE	FALSE	FALSE	FALSE	
403	18	70783	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
404	33	42568	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
405	25	31676	FALSE	TRUE	12	FALSE	TRUE	FALSE	FALSE	FALSE	
406	31	111769	TRUE	FALSE	12	FALSE	TRUE	FALSE	FALSE	FALSE	
407	51	30250	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
408	18	10983	TRUE	FALSE	10	TRUE	FALSE	FALSE	FALSE	FALSE	
409	44	70457	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
410	29	60272	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
411	20	11881	FALSE	TRUE	12	FALSE	TRUE	FALSE	FALSE	FALSE	
412	24	50777	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
413	31	40970	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
414	30	0	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
415	24	32277	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
416	20	102780	FALSE	TRUE	12	FALSE	TRUE	FALSE	FALSE	FALSE	
417	52	50449	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
418	32	122568	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
419	31	50170	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
420	31	22570	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
421	40	120561	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
422	51	81649	TRUE	FALSE	12	TRUE	TRUE	FALSE	FALSE	FALSE	
423	37	110263	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
424	43	121559	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
425	28	102972	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
426	18	22883	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
427	23	52378	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
428	32	70169	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
429	47	81354	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
430	20	81881	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
431	60	51541	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
432	50	81451	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
433	19	101381	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
434	29	112771	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
436	48	90952	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
437	46	72555	TRUE	FALSE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
438	50	71651	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
439	45	52756	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
435	33	100267	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
440	65	81536	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
441	20	91580	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
442	25	61276	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
443	44	102256	FALSE	TRUE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
444	22	80279	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
445	42	110558	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
446	27	92573	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
447	39	32762	TRUE	FALSE	20	TRUE	FALSE	FALSE	FALSE	FALSE	

448	18	22483	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
449	18	122782	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
450	52	80849	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
451	37	31964	FALSE	TRUE	16	FALSE	FALSE	FALSE	FALSE	FALSE	
452	26	80575	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
453	50	40551	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
454	26	20675	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
455	50	22551	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
456	30	60871	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
457	44	22857	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
458	44	92856	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
459	24	80277	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
460	44	12456	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
461	38	60163	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
462	28	41573	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
463	48	60153	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
464	40	102960	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
465	41	92859	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
466	46	111454	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
467	51	111549	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
468	31	20170	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
469	32	40468	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
470	44	51557	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
471	45	71856	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
472	26	83001	FALSE	TRUE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
473	51	52650	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
474	45	111555	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
475	49	102151	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
476	54	11347	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
477	40	30461	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
478	40	20561	TRUE	FALSE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
479	46	12355	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
480	29	111071	TRUE	FALSE	10	TRUE	FALSE	FALSE	FALSE	FALSE	
481	31		TRUE	FALSE	9	TRUE	FALSE	FALSE	FALSE	FALSE	
482	32	20969	TRUE	FALSE		TRUE	FALSE	FALSE	FALSE	FALSE	
483	40	20861	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
484	42	61054	TRUE	FALSE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
485	36	101165	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	

Savannah, GA

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/Caucasion	Native American/ Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/Latino	Ethnicity (specific)
486	43	22058	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
487	36	22265	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
488	35	41166	TRUE	FALSE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
489	38	91562	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
490	22	80179	FALSE	TRUE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
491	26	11575	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
492	29	62372	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
493	25	32676	TRUE	FALSE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
494	20	82781	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
495	19	90882	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
496	22	111378	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
497	29	61572	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
498	43	102657	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
499	69	112331	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
500	53	11448	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
501	42	81959	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
502	19	20282	FALSE	TRUE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
503	25	41976	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
504	34	91566	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
505	21	42880	FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
506	29	70474	FALSE	TRUE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
507	25	120875	TRUE	FALSE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
508	43	31658	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
509	46	30455	FALSE	TRUE	14.5	TRUE	FALSE	FALSE	FALSE	FALES	
510	35	80366	TRUE	FLASE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
511	35	61666	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
512	45	82056	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
513	43	100457	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
514	48	81653	TRUE	FALSE	24	TRUE	FALSE	FALSE	FALSE	FALSE	
515	20	50781	TRUE	FALSE	12	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
516	21	42780	TRUE	FALSE		TRUE	FALSE	FALSE	FALSE	FALSE	
517	31	20770	FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	

Dallas, TX

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/Caucasion	Native American/Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/ Latino	Ethnicity (specific)
519	32	61169	TRUE	FALSE	4	FALSE	FALSE	FALSE	TRUE	FALSE	
520	22	100678	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
521	54	102246	FALSE	TRUE	13.5	TRUE	FALSE	FALSE	FALSE	FALSE	
522	32	103168	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
523	22	120678	FALSE	TRUE	14	FALSE	FALSE	FALSE	TRUE	FALSE	
524	37	102263	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
525	62	51239	TRUE	FALSE		TRUE	FALSE	FALSE	FALSE	FALSE	
526	39	91562	FALSE	TRUE	14	FALSE	FALSE	FALSE	FALSE	TRUE	
527	32	82569	FALSE	TRUE	12	FALSE	FALSE	FALSE	TRUE	FALSE	
528	33	12868	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
529	58	42243	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
530	37	102463	FALSE	TRUE	15	FALSE	FALSE	FALSE	TRUE	FALSE	
531	65	101735	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
532	21	112679	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
533	26	90775	TRUE	FALSE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
534	53	82148	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
535	60	82341	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
536	37	53164	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
537	39	121861	FALSE	TRUE	13	FALSE	FALSE	FALSE	TRUE	FALSE	
538	18	42983	TRUE	FALSE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
539	64	90537	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
540	38	20863	FALSE	TRUE	13.5	TRUE	FALSE	FALSE	FALSE	FALSE	
541	36	100664	FALSE	TRUE	16	FALSE	FALSE	FALSE	TRUE	FALSE	
542	48		FALSE	TRUE	13	TRUE	FALSE	FALSE	FALSE	FALSE	
543	58	101042	TRUE	FLASE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
544	40	52861	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
545	45		FALSE	TRUE	12	TRUE	FALSE	FALSE	FALSE	FALSE	
546	66	92035	TRUE	FALSE	14.5	TRUE	FALSE	FALSE	FALSE	FALSE	
547	43	30359	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
548	25	10876	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	

Boulder, CO

Speaker ID Number	Age	Date of Birth	Male	Female	Education	White/Caucasion	Native American/Native Alaskan	Asian/Pacific Islander	African-American	Mexican-American/Latino	Ethnicity (specific)
1	28	60473	TRUE	FALSE	18	FALSE	FALSE	TRUE	FALSE	FALSE	
2	29	91071	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
3	24	71076	TRUE	FALSE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
4	31	62869	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
5	27	111573	FALSE	TRUE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
6	25	10776	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
7	24	70576	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
8	29	91071	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
9	49	91251	FALSE	TRUE	19	TRUE	FALSE	FALSE	FALSE	FALSE	
10	24	101576	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
11	66	81734	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
12	27	100573	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
13	70	120430	TRUE	FALSE	20	TRUE	FALSE	FALSE	FALSE	FALSE	
14	41	32160	FALSE	TRUE	16	FALSE	FALSE	FALSE	FALSE	TRUE	
15	20	30681	TRUE	FALSE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
16	22	10279	FALSE	TRUE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
17	46	111154	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
18	31	72369	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
19	21	22280	TRUE	FALSE	15	TRUE	FALSE	FALSE	FALSE	FALSE	
20	29	92971	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
21	50	110750	FALSE	TRUE	17	TRUE	FALSE	FALSE	FALSE	FALSE	
22	29	62771	FALSE	TRUE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
23	46	110354	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
24	61	40340	FALSE	TRUE	14	TRUE	FALSE	FALSE	FALSE	FALSE	
25	45	91254	TRUE	FALSE	16	FALSE	FALSE	FALSE	FALSE	TRUE	
26	72	70728	FALSE	TRUE	18	TRUE	FALSE	FALSE	FALSE	FALSE	
27	49	53051	TRUE	FALSE	16	TRUE	FALSE	FALSE	FALSE	FALSE	
28	21	62379	TRUE	FALSE	16	FALSE	FALSE	TRUE	FALSE	FALSE	

CU-Move Project Participants:

The CU-Move project has received significant support from DARPA under the Communicator Program. We have also received support from a number of industry sponsors including HRL and Motorola, which has been instrumental during the data collection efforts. Below, we summarize the personnel involved with the CU-Move project (space limitations prevent listing all the contributions from the CSLR team over the past 2 years).

John H.L. Hansen	CU-Move Principal Investigator
Wayne H. Ward	(dialog research, speech recognition)
Bryan Pellom	(dialog research, speech synthesis, speech recognition)
Xianxian Zhang	(corpus processing and organization)
Murat Akbacak	(noise analysis and modeling)
Mandar Rahurkar	(corpus transcription support)
Jay Plucienkowski	(speech enhancement, beamforming research, hardware)
Stephen Gallant	(microphone array development, corpus hardware support)
Ruhi Sarikaya	(speech recognizer model adaptation - environmental noise)
Umit Yapanel	(speech recognizer model adaptation - environmental noise)
Pongtep Angkititrakul	(corpus development – prompts, display, hardware)
David Cole	(field data collection expert; corpus speaker development
Mieke A. Schierer	(field data collection expert, corpus transcription)
Keith Corson	(lead transcription expert; route dialog Wizard-of-Oz)
Linda Corson	(corpus transcription guidelines supervisor)
Taylor Struemph	(corpus transcription)
Tamara Grivicic	(corpus transcription spell checking/verification)
Elizabeth Elder	(route dialog Wizard-of-Oz monitor)
Carolyn Foster	(route dialog Wizard-of-Oz monitor)
George Figgs	(route dialog Wizard-of-Oz monitor)

Release History:

Release 1.0A Jan. 22, 2002

Contains 153 speakers, Minn., MN

Release 1.1A March 13, 2002

Update includes additional Cell Phone Data that was collected simultaneously at CSLR during WOZ dialog portion. Also includes Route Scenarios from Minn., MN (shows to-and-from destinations used for travel in that city).

Release 2.0A November 8, 2002

Update includes train, dev-test, test set information, extensive speaker information, data from St. Louis, Manchester, Boulder, Dallas, Savannah, and noise analysis results of vehicle.

Release 2.0B March 4, 2005

Includes copies of paper reprints which have employed the CU-Move corpus, as well as a summary of those groups which have purchase the CU-Move Corpus License.

Groups which have the CU-Move Corpus:

- Siemens: CU-Move In-Vehicle Speech Corpus, 2005
- IBM T.J. Watson Research Center: CU-Move In-Vehicle Speech Corpus, 2004
- Motorola Corporation: Wireless Research Group: CU-Move In-Vehicle Speech Corpus, 2001-3
- Panasonic/STL: CU-Move In-Vehicle Speech Corpus, 2002
- Mishubishi (MERL): CU-Move In-Vehicle Speech Corpus, 2002
- Infinitive Speech Systems: CU-Move In-Vehicle Speech Corpus, 2002
- Voice Signal Technologies: CU-Move In-Vehicle Speech Corpus, 2002
- HRL Laboratory : CSLR Center Membership – CU-Move In-Vehicle Research, 2001-2002
- SpeechWorks: CU-Move In-Vehicle Speech Corpus, 2001

Publications (from the Robust Speech Processing Group – CSLR):

Books/Book Chapters:

[1] H. Abut, J.H.L. Hansen, K. Tekeda, *DSP for In-Vehicle and Mobile Systems*, Springer-Verlag Publishing, Oct. 2004.

[2] J.H.L. Hansen, X.X. Zhang, M. Akbacak, U.H.. Yapanel, B.Pellom, W. Ward, P. Angkititrakul, "CU-MOVE: Advanced In-Vehicle Speech Systems for Route Navigation," Chapter 2 in *DSP for In-Vehicle and Mobile Systems*, Springer-Verlag Publishers, 2004.

Journal Papers:

[3] X. Zhang, J.H.L. Hansen, "CSA-BF: A Constrained Switched Adaptive Beamformer for Speech Enhancement and Recognition in Real Car Environments," IEEE Trans. Speech & Audio Proc., vol. 11, no. 6, pp. 733-745, Nov. 2003.

Conference Papers:

[4] X.X. Zhang, J.H.L. Hansen, K. Arehart, J. Rossi-Katz, "In-Vehicle Based Speech Processing for Hearing Impaired Subjects," Interspeech-2004/ICSLP-2004: Inter. Conf. Spoken Language Processing, pp. WeA1101o.3(1-4), Jeju Island, South Korea, Oct. 2004.

[5] X.X. Zhang, K. Takeda, J.H.L. Hansen, T. Maeno, "Audio-Visual Speaker Localization for Car Navigation Systems," Interspeech-2004/ICSLP-2004: Inter. Conf. Spoken Language Processing, pp. Spec3603p.4(1-4), Jeju Island, South Korea, Oct. 2004.

[6] X.X. Zhang, J.H.L. Hansen, K. Arehart, "Speech Enhancement based on a Combined Multi-Channel Array with Constrained Iterative and Auditory Masked Processing," IEEE ICASSP-2004: Inter. Conf. on Acoustics, Speech, and Signal Processing, vol. 1, pp. 229-232, Montreal, Canada, May 2004
[7] X. Zhang, J.H.L. Hansen, "CFA-BF: A Novel Combined Fixed/Adaptive Beamforming for Robust Speech Recognition in Real Car Environments," INTERSPEECH-2003/Eurospeech-2003, pp.1289-1292, Geneva, Switzerland, Sept. 2003. [Xianxian Zhang - Awarded Best Student Paper for Interspeech-2003/Eurospeech-2003 Conference]
[8] U. Yapanel, J.H.L. Hansen, "A New Perspective on Feature Extraction for Robust In-Vehicle Speech Recognition (PMVDR)," INTERSPEECH-2003/Eurospeech-2003, pp.1281-1284, Geneva, Switzerland, Sept. 2003.
[9] M. Akbacak, J.H.L. Hansen, "ENVIRONMENTAL SNIFFING: Robust Digit Recognition for an In-Vehicle Environment," INTERSPEECH-2003/Eurospeech-2003, pp.2177-2180, Geneva, Switzerland, Sept. 2003.
[10] X. Zhang, J.H.L. Hansen, "CSA-BF: Novel Constrained Switched Adaptive Beamforming for Speech Enhancement & Recognition in Real Car Environments," IEEE ICASSP-2003: Inter. Conf. Acoust. Speech & Signal Proc., vol. 2, pp. 125-128, Hong Kong, China, April 2003.
[11] M. Akbacak, J.H.L. Hansen, "Environmental Sniffing: Noise Knowledge Estimation for Robust Speech Systems," IEEE ICASSP-2003: Inter. Conf. Acoust. Speech & Signal Proc., vol. 2, pp. 113-116, Hong Kong, China, April 2003.
[12] J.H.L. Hansen, X. Zhang, M. Akbacak ,U. Yapanel, B. Pellom, W. Ward, "CU-Move: Advances in In-Vehicle Speech Systems for Route Navigation," IEEE Workshop in DSP in Mobile and Vehicular Systems, paper 6.5 (pp. 1-6), Nagoya, Japan, April 4-5, 2003.
[13] U. Yapanel, X. Zhang, J.H.L. Hansen, "High Performance Digit Recognition In Real Car Environments," *ICSLP-2002: Inter. Conf. on Spoken Language Processing*, vol. 2, pp. 793-796, Denver, CO, Sept. 2002.
[14] J. Plucienkowski, J.H.L. Hansen, P. Angkititrakul, "Combined Front-End Signal Processing for In-Vehicle Speech Systems," *Eurospeech-2001*, vol. 3, pp. 1573-1576, Aalborg, Denmark, Sept. 2001.
[15] J.H.L. Hansen, P. Angkititrakul, J. Plucienkowski, S. Gallant, U. Yapanel, B. Pellom, W. Ward, "CU-Move : Analysis & Corpus Development for Interactive In-Vehicle Speech Systems," *Eurospeech-2001*, vol. 3, pp. 2023-2026, Aalborg, Denmark, Sept. 2001.
[16] J.H.L. Hansen, J. Plucienkowski, S. Gallant, B.L. Pellom, W. Ward, "CU-Move: Robust Speech Processing for In-Vehicle Speech Systems," *ICSLP-2000: Inter. Conf. Spoken Language Processing*, vol. 1, pp. 524-527, Beijing, China, Oct. 2000.

Copies of these papers are available at:
http://cslr.colorado.edu/people/jhlh/JHLH-resume-Mar05.htm
or you can contact John H.L. Hansen at: John.Hansen@colorado.edu

Xianxian Zhang and John H.L. Hansen, "CFA-BF: A Novel Combined Fixed/Adaptive Beamforming for Robust Speech Recognition in Real Car Environments," Eurospeech-2003, pp. 1289-1292, Switzerland, Sept. 2003.

CFA-BF: A Novel Combined Fixed/Adaptive Beamforming for Robust Speech Recognition in Real Car Environments

Xianxian Zhang and John H.L. Hansen,

Robust Speech Processing Group
Center for Spoken Language Research
University of Colorado Boulder, Campus Box 594
(Express Mail: 3215 Marine Street, Room E-265)
Boulder, Colorado 80309-0594
303 – 735 –5148 (Phone) 303 – 735 – 5072 (Fax)
http://cslr.colorado.edu/
{yapanel, jhlh, sarikaya, bp}@cslr.colorado.edu, (email)

EUROSPEECH-2003 / INTERSPEECH-2003
Switzerland, Sept. 1-4, 2003.

International
Speech
Communication
Association

CFA-BF: A Novel Combined Fixed/Adaptive Beamforming for Robust Speech Recognition In Real Car Environments

Xianxian Zhang and John H. L. Hansen +

Robust Speech Processing Group, Center for Spoken Language Research
University of Colorado at Boulder, USA
xianxian.zhang@colorado.edu john.hansen@colorado.edu

Abstract

Among a number of studies which have investigated various speech enhancement and processing schemes for in-vehicle speech systems, the delay-and-sum beamforming (DASB) and adaptive beamforming are two typical methods that both have their advantages and disadvantages. In this paper, we propose a novel combined fixed/adaptive beamforming solution (CFA-BF) based on previous work for speech enhancement and recognition in real moving car environments, which seeks to take advantage of both methods. The working scheme of CFA-BF consists of two steps: source location calibration and target signal enhancement. The first step is to pre-record the transfer functions between speaker and microphone array from different potential source positions using adaptive beamforming under quiet environments; and the second step is to use this pre-recorded information to enhance the desired speech when the car is running on the road. An evaluation using extensive actual car speech data from the CU-Move Corpus shows that the method can decrease WER for speech recognition by up to 30% over a single channel scenario.

1. Introduction

The increased use of mobile telephones in cars has created a greater demand for hands-free, in-car installations. Many countries now restrict the use of hand-held cellular technology while operating a vehicle. As such, there is a greater need to have reliable voice capture within automobile environments. However, the distance between a hands-free car microphone and the speaker will cause a severe loss in speech quality due to changing acoustic environments. Therefore, the topic of capturing clean and distortion-free speech under distant talker conditions in noisy car environments has attracted much attention. Microphone array processing and beamforming is one promising area which can yield effective performance.

A number of beamforming methods which are suitable for speech enhancement and recognition in car environments have been proposed in the past. In a study by Nordholm, *et. al* [1], they formulate a simple built-in calibration procedure for data collection instrumentation in the car environment. Their working scheme is to find the transfer function among the speaker, jammer signal, and microphone array in a quiet setting, and assume this function does not change when the car is moving on the road. This algorithm is one of several typical beamforming algorithms that have been used in car environments. However, from our analysis using real car data we collected, we found that different drivers position their heads differently, and this will result in deviations in the transfer function between the speaker and microphone. In another study, Compernolle [2] presented an approach using switching adaptive filters, with no a priori knowledge about the speech source. The filters have two sections, where the first section implements an adaptive look direction and cues in on the desired speech, while the second section acts as a multi-channel adaptive noise canceller. This method is able to simultaneously track the movement of the speaker source and compensate for the transfer function between the microphone array and speaker in real-time. This study represents an important step forward for in-car speech applications. Another study by Oh, *et. al* [3] applied a Griffiths-Jim beamformer in a car environment with a 7-channel microphone array. They evaluated Signal-to-noise ratio (SNR) and word-error-rate (WER) improvement of their algorithm, and compared this to the case when only Delay-and-sum beamforming (DASB) is used. Their general recommendations were that the generalized side-lobe canceller (GSC) was relatively stable and robust. However, from our analysis using real car data we collected, we found that noise signals with high frequency energy, such as road bump noise, which routinely happens for road surface repairs of potholes or expansion joints across bridges, will make the GSC unstable. Zhang and Hansen [4] proposed a method to identify this kind of noise and thereby allow the adaptive filters to work more robustly. In a study by Shinde, Takeda and Itakura [5], they presented a multichannel method for noisy speech recognition which estimates the log spectrum of speech for a close-talking microphone based on a multiple regression of the log spectra (MRLS) of noisy signals captured by the distributed microphones. This method was reported to improve speech recognition performance by up to 20%, but it requires a specific microphone arrangement in the car. It should also be noted that the noise signals captured by distributed microphones within the car are not necessarily the real noise that reaches the close-talking microphone. Visser, Otsuka and Lee [6], presented a speech enhancement scheme which combined a spatial and temporal processing strategy to handle reverberation, highly interfering sources and background noise without the need of microphone arrays nor a priori speech or noise models. These methods were reported to have good performance under a single controlled driving condition (i.e., windows closed traveling at a given speed).

While a number of studies have investigated various speech enhancement and processing schemes for in-vehicle speech systems, the majority of results are conducted under controlled simulated conditions inside a room or with pre-recorded car noise. Little research has been performed using actual voice data collected in the car with associated environmental noise conditions. Because of the variety of simulated in-vehicle evaluation scenarios, it is difficult to compare performance between studies, and to predict if simulated performance will hold for actual, in-vehicle conditions. In Zhang and Hansen [4], an analysis was performed on data recorded in various car noise environments from across the United States. There, we proposed a

This work was supported by DARPA through SPAWAR under Grant No. N66001-8906

constrained switched adaptive beamforming algorithm (CSA-BF), which detects the head movement of the driver and adjusts the time delay between microphones automatically. That method was shown to decrease WER (Word Error Rate) for speech recognition by up to 31% and improve speech quality by up to 5.5dB on the average simultaneously, using the CU-Move corpus [7].

In this paper, we first analyze potential driver movement during voice interaction by selecting ten-speakers from the CU-Move corpus [7,8], then propose a combined fixed/adaptive beamforming (CFA-BF) scheme designed specially for robust speech recognition in car noise environments. Our proposed method combines fixed and adaptive beamforming and also applies source localization techniques. Therefore, it has the following advantages.

- Low computational complexity with robustness;
- Target signal distortion reduction by omitting the parameter adjustment in adaptive filters;
- Automatically tracking driver movement, and no speech range definition is needed;
- Directional sources can be suppressed;
- Especially suitable for use in car noise environments.

2. Prior CSA-BF Beamforming Algorithm

The CSA-BF algorithm consists of a speech/noise constraint section (CS), a speech adaptive beamformer (SA-BF), and a noise adaptive beamformer (NA-BF). A speech range is defined and judged in the constraint section. The desired speech signal is enhanced by the speech beamformer, and noise is suppressed by the noise beamformer. A set of adaptive filters are used to perform the beam steering. Also, a normalized LMS algorithm is used to update the filter coefficients. The most novel advantage of CSA-BF method is that source movement can be tracked and directional sources can be suppressed with reduced target signal distortion. However, the adaptive filters used in CSA-BF increase the computation complexity greatly, which limits the implementation of CSA-BF algorithm in real-time. Another disadvantage of CSA-BF is the sensitivity of parameter setting for the adaptive filters. From the experiment results in [4], we know that if the optimal parameter settings for CSA-BF are altered slightly, the WER degrades slightly because of speech leakage. A comparison with DASB was also performed. Compared with adaptive beamforming, the computation complexity of DASB is quite low. However, DASB will lose accuracy in estimating the time delay if the driver's physical position changes significantly.

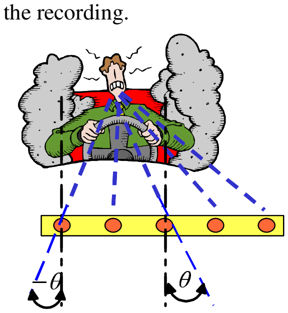
3. Source Location Analysis In Real Car Environments

In order to analyze the movement of the driver's head in the car during voice interaction, we selected 10 speakers from the CU-Move database [7, 8] that are balanced across gender and age. Next we use the TEO criterion described in [4] to decide the speech period for each of the 10 speakers and apply the adaptive LMS filter technique [9] to locate the position of the head of each speaker (source). Table 1 shows the entire recording time for each speaker and the percentage (%) of time each speaker's head stays in a certain position.

Position No	Speaker Number									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	10
	Amount of time in minutes for recording									
	5.6	8.2	7.4	8.1	8.2	7.4	6.5	6.1	6.6	6.4
0	**39**	**57**	1			8	14		2	
1	**36**					**50**	**82**		**80**	
2	4	9			5	17	2		8	**55**
3		.1			14	18		**67**	3	38
4					.5	4			.4	
5		2			.2	.4			.5	
6		.3		1	.3				.1	
7				**66**		.2				
8				10		.4				
-1	2		**91**					1		
-2			8							
-8				1						
unknown	19	32	0	22	**80**	3	2	32	6	7

Table 1: Percentage Time Of Each Speaker Source Location Over CU-Move In-vehicle Recording
(i.e., Speaker 1 spends 39% of his total 5.6 minutes of speech in digits portion with head position 0 from Fig. 1)

Fig. 1 shows the position number in Table 1 corresponding to the source angle to the axis of the microphone array during the recording.



position number	corresponding θ
0	0
1	2.8°
2	5.6°
3	8.4°
4	11.2°
5	14°
6	16.8°
7	19.6°
8	22.4°
-1	-2.8°
-2	-5.6°
unknown	unknown

Fig. 1: Relation Between Position Number and Angle of Source

From this table, we find that even during no more than 9 minutes of voice recording, the driver will change his head position often. Fortunately, for each speaker, we can always find a dominant position. The reason we have some unknown positions is that the source location technique we employ at times cannot make a reliable decision as to the current source location. This may happen when the noise level is very high, the noise changes too fast, and/or the step-size of the filter is too large. This is actually a common situation for in-vehicle systems because of the complex noise situations, and the limitations of the adaptive LMS filter technique. This also is a motivation for the proposed CFA-BF algorithm.

4. CFA-BF: Combined Fixed/Adaptive Beamforming

In this proposed method, we assume that if the source position (driver's head) does not change, then the transfer function between the speaker and microphone array in a quiet setting will not change if the car is moving on the road. So, we find the transfer function between the speaker and microphone array for different possible source positions when the car is in a quiet environment (for example, parking plot), and pre-store them for later use when the car is driven on the road.

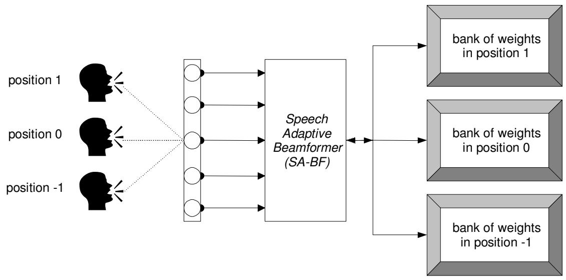


Fig. 2: Working Scheme for the Proposed CFA-BF

4.1. Source Location Calibration – Adaptive Beamforming

As is well known, an adaptive algorithm such as normalized Least Mean Square algorithm (NLMS) can more easily to reach its convergence behavior in quiet or stationary noise environments, than under non-stationary noise environments (for example, car noise environments). Also, from source location analysis of CU-Move Corpus in Sec. 3, we know that although different drivers will move their heads in different positions, almost all of them keep themselves in one position more than 50% of the time. Thus, we can study possible positions which the driver's head can reach inside a car, and then apply the previous developed CSA-BF [4] to pre-record the weight coefficients of the adaptive filters in speech adaptive beamforming (SA-BF) from all the possible source positions in a quiet environment. Fig. 2 is the working scheme of the source calibration procedure. Here, we only show 3 positions. A normalized LMS algorithm is used to update the filter coefficients, and the update equations are given as follows:

$$e_{1i}(n) = \mathbf{x}_1(n - L/2) - \mathbf{w}_{1i}^T(n)\mathbf{x}_i(n) \qquad (1)$$

$$\mathbf{w}_{1i}(n+1) = \mathbf{w}_{1i}(n) + \frac{2\mu}{\mathbf{x}_i^T(n)\mathbf{x}_i(n)} e_{1i}(n)\mathbf{x}_i(n) \qquad (2)$$

The coefficients stored in the bank of weights impliment the transfer functions between the microphone array and the speaker in different positions respectively. They also reflect the relative delays between microphones. Fig. 3 shows how the SA-BF operates.

4.2. Target Signal Enhancement – Fixed Beamforming

Fig. 4 shows the working scheme of the target speech enhancement. At this point, we have the transfer functions from the speaker in different positions, (i.e., weight coefficients ($[W_{12}^\theta, W_{13}^\theta, W_{14}^\theta, W_{15}^\theta]$).With the help of a source localization technique, we find the source position first and then extract the corresponding weight coefficient bank from the pre-recorded weights and use them in this section.

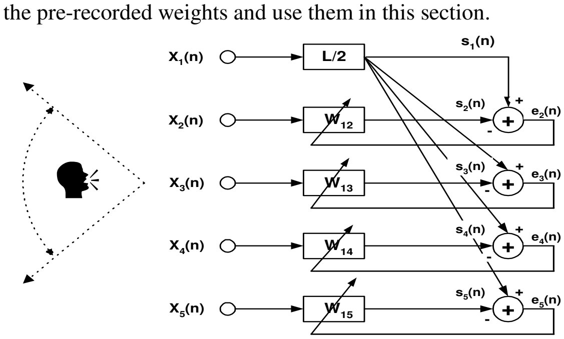


Fig 3: Structure of Speech Adaptive Beamformer (SA-BF)

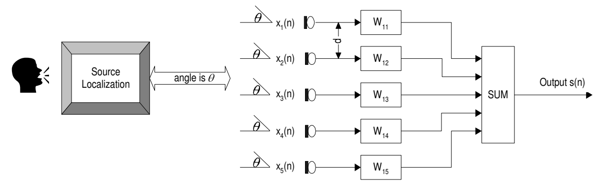


Fig. 4: Structure of Fixed Beamforming for Target Signal Enhancement

With this procedure, the enhanced speech will be given as follows:

$$s(n) = \sum_{i=1}^{5} \mathbf{w}_{1i}^T(n)\mathbf{x}_{1i}(n) \qquad (3)$$

where $[W_{12}, W_{13}, W_{14}, W_{15}]$ are functions of the angle between the source and axis of the microphone array θ, and W_{11} is a pure delay which is half of the filter length (i.e., L/2).

5. Performance Evaluation

5.1. CU-Move Corpus

The CU-Move [8] database includes 5 parts: command and control words, digit strings of telephone and credit card numbers, street names and addresses, phonetically balanced sentences, and Wizard of Oz interactive navigation conversation. A total of 500 speakers, balanced across gender and age, produced over 600GB of data during a six-month collection effort across the United States. The database and noise conditions are discussed in detail in [7]. We point out that the noise conditions are changing with time and are quite different in terms of SNR, stationarity and spectral structure. In this study, we chose 10 speakers from approximately 100 speakers in Minn., MN (i.e., Release 1.1A) and use the digits portion that includes speech under a range of varying complex car noise environments and contains approximately 40 words.

5.2. Experiment Establishment

For the proposed CFA-BF algorithm, careful calibration of the weight coefficients and the source location decision have significant impact on the performance of the algorithm. In order to evaluate the performance of the CFA-BF under the non-ideal calibration and source location process, we establish experiments as follows:

- Use CSA-BF to process each of the ten speakers respectively; the constraint we use here is the TEO criterion described in [1] only;
- Store the weight coefficient set of the speech beamformer (SA-BF) which has the dominant source position, and choose the best from this set as the calibrated weight set for SA-BF for this speaker.
- Use the calibrated weight set to re-process the data for this speaker (i.e., delay-and-sum).

If CFA-BF can perform better result DASB and SA-BF under this established experiment, it will operate much better than in ideal experimental conditions. We have shown in [4] that with noise cancellation processing activated, both SEGSNR and WER results can be improved compared with SA-BF. In this study, we disable the cancellation processor, since if the speech quality (i.e., one of the outputs of SA-BF, which is used as the reference for noise cancellation processor) is improved, the output of GSC will also be improved.

5.3. Evaluations

For evaluation, we consider two different performance measures using CU-Move data. One measure is the Segmental Signal-to-Noise Ratio (SEGSNR)[10] which represents a noise reduction criterion for voice communications. The second performance measure is Word Error Rate (WER) reduction, which reflects benefits for speech recognition applications. The Sonic Recognizer [11] is used to investigate speech recognition performance. Since the size of the processed data is not large enough for recognizer evaluation, therefore, we adopted the cross-validation method from [12].

5.4. Experiments Results

Fig. 5 shows the SEGSNR results for reference single channel3, DASB, SA-BF, and proposed CFA-BF.

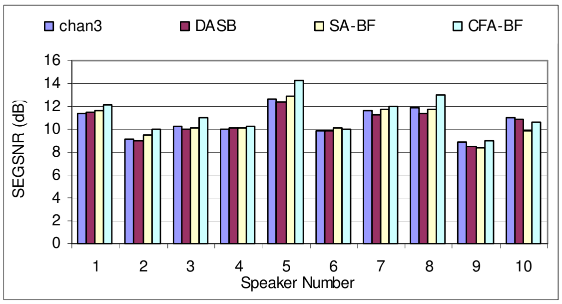


Fig. 5: SEGSNR Performance for Ref. 3 Microphone and different Beamforming Scenarios

Table 2 shows average SEGSNR improvement, average WER (Word Error Rate), CORR (Word Correct Rate), SUB (Word Substitution Rate), DEL (Word Deletion Rate) and INS (Word Insertion Rate) for the 10 speakers.

method / measure	Chan3	DASB	SA-BF	CFA-BF
Ave. (dB) SEGSNR	10.77	10.58	10.7	11.34
WER	**10.71**	8.28	7.98	**7.51**
SUB	4.76	3.9	3.76	3.51
DEL	4.75	2.35	3.88	2.19
INS	3	3.11	3.16	2.96
CORR	92.28	94.83	95.19	95.46

Table 2: Average SEGSNR, WER, CORR, SUB, DEL and INS for Ref. 3 Microphone and Beamforming Scenarios

Fig. 6 illustrate average SEGSNR improvement and WER speech recognition performance results respectively. The average SEGSNR results are indicated by the bars using the left-side vertical scale (dB), and the WER improvement is the solid line using the right-side scale (%).

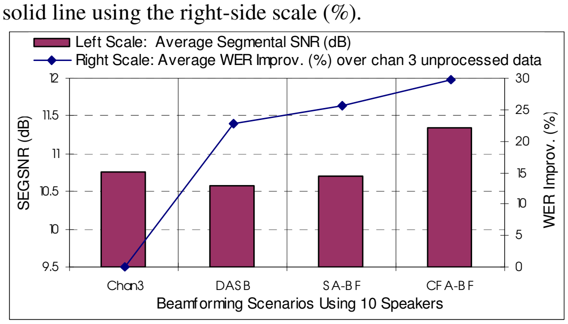


Fig. 6: SEGSNR and WER Results for Ref. 3 Microphone and Beamforming Scenarios Using CU-Move 10 Speakers

From these results, we make the following observations:

(i.) Employing the proposed combined fixed/adaptive beamforming method, increases SEGSNR slightly, but some variability exists across speakers.

(ii.) However, DASB, SA-BF and the proposed method can provide WER improvement by 22.8%, 25.6% and 29.9% respectively over a single microphone (i.e., channel 3).

6. Conclusions and Future Work

In this paper, we have proposed a novel combined fixed/adaptive beamforming method (CFA-BF) for robust speech recognition in real car environments based on experiments using voice data recorded in moving car environments. We demonstrated that the CFA-BF can improve SEGSNR slightly, and improve speech recognition performance by decreasing WER by 29.3% using CU-Move in-vehicle speech data. We have shown that this method outperforms a single channel microphone (channel 3), traditional delay-and-sum beamforming and our previous speech adaptive beamformer (SA-BF).

However, there remain some issues to address for future work:

- Perform source localization calibration in a quiet environment, such as parking plot, and use a larger portion of the CU-Move Corpus to evaluate the performance of CFA-BF;
- Improve the accuracy of source localization by applying alternative source localization techniques, such as CSP (cross-power spectrum technique), and decrease the percentage of unknown positions.
- Activate the GSC noise canceller processor after signal enhancement, and improve the SEGSNR performance without affecting WER improvement.

7. References

[1] Nordholm *et. al.*, "Adaptive Microphone Array Employing Calibration Signals: Analytical Evaluation," *IEEE Trans. on Speech & Audio Proc.*, vol 7, no. 3, May 1999.

[2] Compernolle, "Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings," in *ICASSP'1990*.

[3] Oh *et. al.*, "Hands-free Voice Communication in an Automobile with a Microphone Array," in *ICASSP'1992*.

[4] Zhang and Hansen, "CSA-BF: Novel Constrained Switched Adaptive Beamforming For Speech Enhancement & Recognition In Real Car Environments," in *ICASSP'2003*.

[5] Shinde *et. al.*, "Multiple Regression of Log-spectra for In-Car Speech Recognition," in *ICSLP'2002*.

[6] Visser *et. al.*, , "A Spatio-Temporal Speech Enhancement Scheme for Robust Speech Recognition," in *ICSLP'2002*.

[7] Hansen *et. al.*, "CU-Move: Analysis & Corpus Development for Interactive In-vehicle Speech Systems," in *Eurospeech'01*.

[8] http://cumove.colorado.edu

[9] Reed *et. al.* "Time Delay Estimation Using the LMS Adaptive Filter-Static Behavior", *IEEE Trans. On Acous., Speech and Signal Proc.,* June 1981.

[10] http://www.nist.gov

[11] Pellom, "Sonic: The University of Colorado Continuous Speech Recognizer", *University of Colorado, Technical Report #TR-CSLR-2003-01*, Boulder, Colorado.

[12] Rabiner and Juang, "*Fundamentals of Speech Recognition*", Englewood Cliffs, NJ: Prentice-Hall, 1993.

[13] W. Koroljow and G. Gibian, US patent #6,154,552, May, 1998.

M. Akbacak, J.H.L. Hansen, "Environmental Sniffing: Noise Knowledge Estimation for Robust Speech Systems," IEEE ICASSP-2003: Inter. Conf. Acoust. Speech & Signal Proc., vol. 2, pp. 113-116, Hong Kong, China, April 2003.

Environmental Sniffing: Noise Knowledge Estimation for Robust Speech Systems

Murat Akbacak, John H.L. Hansen

Robust Speech Processing Group
Center for Spoken Language Research
University of Colorado Boulder, Campus Box 594
(Express Mail: 3215 Marine Street, Room E-265)
Boulder, Colorado 80309-0594
303 – 735 –5148 (Phone) 303 – 735 – 5072 (Fax)
http://cslr.colorado.edu/
Murat.Akbacak@colorado.edu John.Hansen@colorado.edu, (email)

IEEE ICASSP-2003: Inter. Conf. On Acoustics, Speech, and Signal Processing, Hong Kong, China, April 2003.

ICASSP 2003 Hong Kong
IEEE International Conference on Acoustics, Speech, and Signal Processing

ENVIRONMENTAL SNIFFING: NOISE KNOWLEDGE ESTIMATION FOR ROBUST SPEECH SYSTEMS

Murat Akbacak and John H.L. Hansen

Robust Speech Processing Group, Center for Spoken Language Research
University of Colorado at Boulder, Boulder, CO, 80302, U.S.A
{murat,jhlh}@cslr.colorado.edu   Web: http://cslr.colorado.edu

ABSTRACT

In this paper, we propose a framework for extracting knowledge concerning environmental noise from an input audio sequence and organizing this knowledge for use by other speech systems. To date, most approaches dealing with environmental noise in speech systems are based on assumptions concerning the noise, or differences in collecting and training on a specific noise condition, rather than exploring the nature of the noise. We are interested in constructing a new speech framework entitled *Environmental Sniffing* to detect, classify and track acoustic environmental conditions. The first goal of the framework is to seek out detailed information about the environmental characteristics instead of just detecting environmental changes. The second goal is to organize this knowledge in an effective manner to allow smart decisions to direct other speech systems. Our current framework uses a number of speech processing modules including the Teager Energy Operator (TEO) and a hybrid algorithm with T^2-BIC segmentation, noise language modeling and GMM classification in noise knowledge estimation. We define a new information criterion that incorporates impact of noise into Environmental Sniffing performance. We use an in-vehicle speech and noise environment as a test platform for our evaluations and investigate the integration of Environmental Sniffing into an Automatic Speech Recognition (ASR) engine in this environment. Noise classification experiments show that the hybrid algorithm achieves an error rate of 25.51 % , outperforming a baseline system by a relative 7.08%.

1. INTRODUCTION

Significant advances in speech technology have been achieved in applications where the environmental condition is constant. Most recently, research focus has shifted to the real-world environments where changing environmental conditions represent significant challenges in maintaining speech system performance.

This problem has been taken into consideration especially in ASR applications since the recognition performance degrades substantially due to changes in the environment. One of the first ASR tasks that have changing environmental conditions is for automatic transcription of “Broadcast News” (BN). Several research groups have worked on this task to increase recognition performance. These studies [1, 2] have the underlying goal of training for acoustic conditions that are specific for each system (speech conditions include : F0- prepared, F1- spontaneous,F2- degraded acoustics, F3- music background, F4- noise background, F5- nonnative speakers, and FX- other speech) and directing the ASR engine to a single recognizer for each acoustic condition. The downside of this method is that it tries to model many different kinds of environmental conditions with a single model, with the hope that such a background noise model would be able to capture this huge variability.

Later, as computational power has increased with the help of high-speed computers, a parallel bank of recognizers has been used in a ROVER paradigm for tasks such as Speech In Noisy Environments (SPINE) where many different environmental conditions exist. Different recognizers intentionally employing a range of feature processing or adaptation methods are normal for a ROVER based LVCSR solution. This may involve different features during the feature extraction step, different noise compensation schemes in the enhancement step, or different model adaptation schemes individually or in parallel. Finally, the hypothesis with the highest probability at the output of the decoders is chosen as the final decision of the ROVER. Although significant improvement has been achieved using the ROVER paradigm, it is not optimal in terms of computational performance. It is also highly possible that one recognizer may not have the highest probability at all times during decoding, implying that the selected recognizer may be the best in a global sense but not in a local sense.

To overcome the disadvantages of these methods as well as to have acceptable error rates in ASR systems in changing environmental conditions, we propose a new speech framework called *Environmental Sniffing*. The goal will be to do smart tracking of environmental conditions and direct the ASR engine to use the best local solution specific to each environmental condition. For example, instead of running parallel feature extractors in a ROVER paradigm, the Environmental Sniffing framework will direct the ASR engine to use only one feature extractor which gives the best performance for a specific environmental condition. In this way, we optimize both the computational effort and overall system performance of the ASR.

On the other hand, Environmental Sniffing is also useful for automatic transcription of noise where the accuracy is much lower than that of transcription of speech. Considering the fact that there are no standards for noise transcription in audio material, it is critical to automatically transcribe environmental noise with high accuracy for more effective speech system training.

The organization of our paper is as follows. In Section 2, a general system architecture for Environmental Sniffing is presented. In Section 3, we specialize the general framework for sniffing environmental noise for in-vehicle systems. In Section 4, evaluations of the framework integrated into an in-vehicle ASR engine is presented. Section 5 discusses some further research issues for sniffing. Conclusion is given in Section 6.

This work was supported in part by DARPA under Grant No. N66001-8906 and NSF Cooperative Agreement No. IIS-9817485

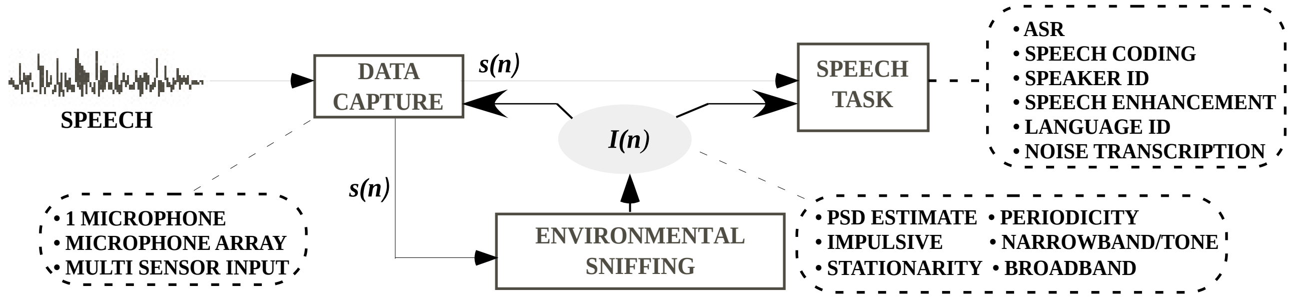


Fig. 1. Environmental Sniffing system architecture diagram.

2. SYSTEM ARCHITECTURE

Environmental Sniffing can be integrated into any speech task having some degree of concern about acoustic environmental conditions. Environmental Sniffing extracts knowledge about the acoustic environmental conditions and passes this knowledge to the speech task. A proposed general system architecture diagram is shown in Fig. 1. Digitized speech is denoted as $s(n)$, captured from an input sensor (i.e., single or multi-microphone) and acoustic environmental noise knowledge as $I(n)$ which is a function of $s(n)$. In a sample scenario, $s(n)$ may be the audio data recorded in a vehicle with a microphone array, the speech task may include model adaptation within an ASR system, and $I(n)$ may consist of the existing noise types with time tags and the power spectral estimates of the environmental noise with a stationarity measure. Here, $I(n)$ may also contain a suggestion to use one of several adaptation schemes (Jacobian adaptation, MLLR, PMC, etc.) which gives the best performance for the environmental noise knowledge estimated through Environmental Sniffing.

3. IN-VEHICLE ENVIRONMENTAL SNIFFING

Within the framework of Environmental Sniffing from Fig. 1, we specialize our solution for an in-vehicle hands-free car navigation environment. The motivation for selecting this environment is the huge diversity of acoustic environmental conditions and the need to maintain near real-time performance for route navigation dialogs [3].

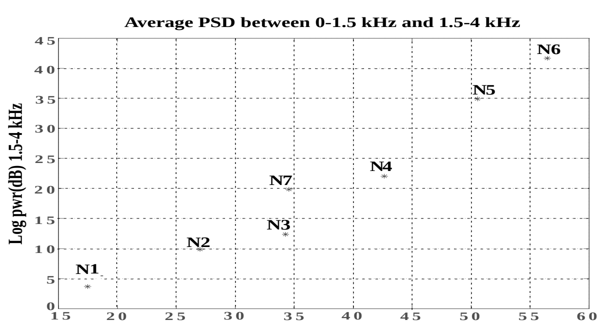


Fig. 2. Scatter plot of low (0-1.5 kHz) versus high (1.5-4 kHz) frequency noise dB-energy content for noises N1 through N7.

Having collected in-vehicle acoustic data (i.e., in a Blazer SUV) using a 17 mile route which contains samples of all driving conditions expected for use in city and rural areas, we identified the primary noise conditions of interest (noise conditions include: N1- idle noise consisting of the engine running with no movement and windows closed, N2- city driving without traffic and windows closed, N3- city driving with traffic and windows closed, N4- highway driving with windows closed, N5- highway driving with windows 2 inches open, N6- highway driving with windows half-way down, N7- windows 2 inches open in city traffic, N0- others), which are considered as long term acoustic environmental conditions. Other acoustic conditions (idle position with air-conditioning on, etc.) are matched to these primary classes having the closest acoustic characteristic. Fig. 2 shows the average power spectrum density for low (0-1.5 kHz) versus high (1.5-4 kHz) frequency energy content of long term noises. The diversity of noise energy content suggests that a single noise model would not be capable of addressing changing noise conditions for a subsequent speech task.

Short term acoustic environmental conditions occurring within long term conditions include TS- turn signal noise, WB- wiper blade noise, TN- tone noise, IM- impulsive noise. These conditions are expected to be present in conjunction with one of the long-term noises.

As shown in Fig. 3, a hybrid method of T^2-BIC segmentation and GMM classification followed by a decision smoothing is used to detect, classify and track long-term noises. T^2-BIC uses Hotelling's T^2-Statistic to pre-rank potential acoustic break points which are evaluated using a Bayesian Information Criterion [4].

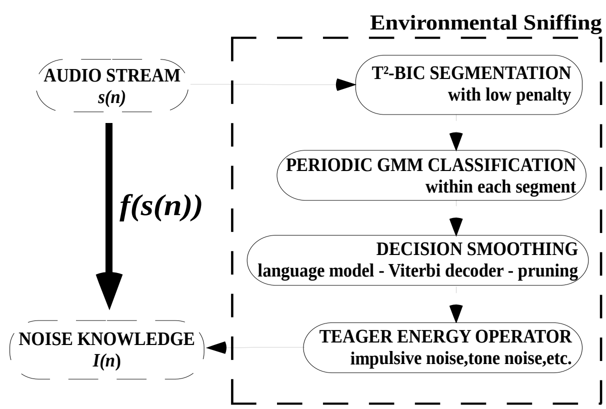


Fig. 3. Flow diagram for Environmental Noise Sniffing.

As Fig. 3 shows, the incoming audio stream is first segmented into acoustically homogeneous speech blocks using our T^2-BIC segmentation scheme with a low false alarm penalty (i.e. false alarms are tolerable to ensure we capture all potential marks, both true and false). Within each segment, GMM classification runs periodically to classify each non-overlapping T-frame-length block.

Decision smoothing is applied to the resulting decision sequence of each segment. This process is similar to Language Modeling, considering the fact that some noise transitions are not possible although they may appear at the output of the GMM classifier. Transition probabilities are generated from training data using bi-gram language modeling with a noise type for each 15-frame word block. Calculated transition probabilities are shown in Fig. 4.

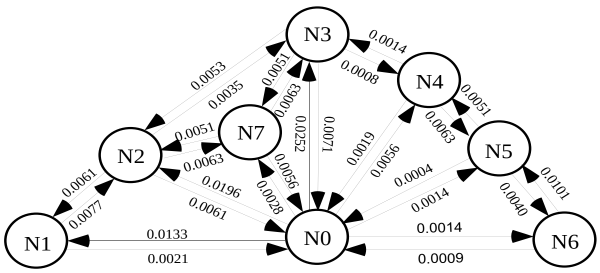


Fig. 4. Noise Language Modeling.

We use Viterbi decoding to find the most likely decision sequence given the classification probability list of each decision region within the segment. Each noise class has an initial probability which is proportional to the number of occurrences within the N-best position at the classifier output along the segment. Noise classes whose number of occurrences within the N-best position is less than a threshold are pruned during decision smoothing. We can formulate this as follows:

$$\alpha_1 n_1 + \alpha_2 n_2 + \ldots + (\alpha)_N n_N \leq \gamma \qquad (1)$$

where n_i is the occurrence number in the i^{th} position in the score list, α_i is the corresponding weight coefficient and γ is the threshold.

Since our Environmental Sniffing framework is not a speech system itself and works with other speech systems, noise knowledge detection performance for each noise type (P_i) should be weighted by a coefficient which is determined by the importance that noise type plays in the speech application with Environmental Sniffing (i.e., if noise impacts the speech task performance significantly, impact coefficient I is set high). For in-vehicle ASR, these coefficients $(I_1, I_2, \ldots, I_n)$ will reflect the impact each noise type has on WER. We can formulate this as follows:

$$Critical\ performance\ rate \triangleq \sum_{i=1}^{n} I_i P_i \quad \sum_{i=1}^{n} I_i = 1 \qquad (2)$$

With this performance rate measure, the potential output score can range from 0-100 if P_i is a classification rate, or 0-1 if P_i is a probability.

4. EVALUATIONS

We evaluate the performance of our framework using an in-vehicle noise database of 3 hours collected in 6 experimental runs using the same route and the same vehicle on different days and hours. A microphone array and 8-channel digital recorder previously used for CU-Move in-vehicle speech data collection were employed [3]. The database does not contain speech. Fifteen noise classes are transcribed during the data collection by a transcriber sitting in the car. The time tags are generated instantly by the transcriber. After data collection, some noise conditions are grouped together, resulting in 8 acoustically distinguishable noise classes as listed in Sec. 3. For each noise class, a 4-mixture GMM is trained using 2.5 hours of data. We use 12 dimensional MFCC feature vectors during our evaluations. In both training and test data, half of the time, long-term and short-term noise conditions are approximately equally balanced across time.

4.1. Long Term Noise

First, we test long-term noise classification error performance by running the classifier periodically with a period of 15 frames without segmenting the test data. Fig. 5 shows noise classification error performance by selecting the most likely model (solid bar to left in each pair) [avg. 34.73% error], and using the two highest probable models (cross-hatch bar to right in each pair) [avg. 13.23% error]. Some noise types (N4-highway driving, windows closed) are significantly affected by selecting the top two models out of 8 in the noise space.

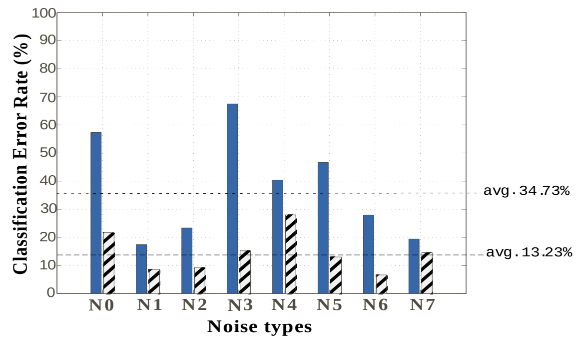


Fig. 5. Classification error performance of having the correct noise type in the first position (1^{st} bar in each set) and first two positions (2^{nd} bar in each set).

In our "Classical classification algorithm" for noise, a segment of data is scored once. As shown in Fig. 3, our "Hybrid Algorithm" has periodical classifications within a segment and subsequently smoothes the final decision sequence using the language model and pruning.

Next, we segment the noise test data using T^2-BIC with different false alarm penalties ($\lambda = \{0.3, 0.4, 0.5, 0.6\}$). During decision smoothing in the hybrid algorithm, we use the values $N = 2$, $\alpha_1 = 0.7$, $\alpha_2 = 0.3$, and pruning threshold $\gamma = 0.7$. Fig. 6 shows error rates for both methods. You can see that classical method is worse than the hybrid algorithm in terms of classification performance even if the hand label segmentation is provided.

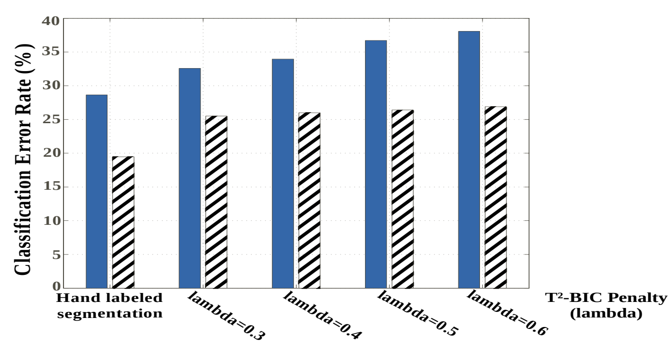


Fig. 6. Error performance of the classical method (1^{st} bar in each set) and the hybrid algorithm (2^{nd} bar in each set) with hand label segmentation, and a range of lambda values for T^2-BIC segmentation.

To calculate the overall performance using Eqn. 2, we ran speech recognition tests using CSLR's Large Vocabulary Continuous Speech Recognizer SONIC [5] on the TI-DIGITS database af-

Degrading noise	N01	N02	N03	N04	N05	N06	N07
WER	1.1%	2.3%	2.7%	4.1%	8.1%	8.5%	3.7%
I-measure	0.04	0.08	0.09	0.13	0.27	0.28	0.11

Table 1. Speech Recognition Tests.

ter degrading the clean speech with our noise types at 10 dB SNR. Models trained from clean speech were used for testing. WER results are shown in Table 1 as well as the impact I-measures of each noise type.

I_i's are assigned proportionally to WER's and they sum to one. Using Eqn. 2 with these values, we found the critical performance rate to be 65.41% for the classical classification method and 69.61% for the hybrid algorithm using 0.3 as the penalty parameter for T^2-BIC.

4.2. Discussion on Detecting Short Term Noise

We have the following assumptions about the human auditory system: hearing is the process of detecting energy at a particular frequency and the human auditory system is assumed to be a filtering process which partitions the entire audible frequency range into many critical bands. These assumptions provide motivation for use of the Teager Energy Operator (TEO) [6], to detect impulsive noise, tone noise and periodic noise observed in the in-vehicle environment since they appear as sudden energy changes and occupy a certain frequency band. What distinguishes these energy changes from those appearing during speech is that they do not have an observed modulation scheme like speech. Using this knowledge, we can automatically detect short-term noises within noisy-speech. Fig. 7 gives an idea of how TEO processing works for turn signal which occupies narrow time slots and a wide frequency band. The last figure (Fig. 7-c) clearly shows detection locations where turn signal noise is present.

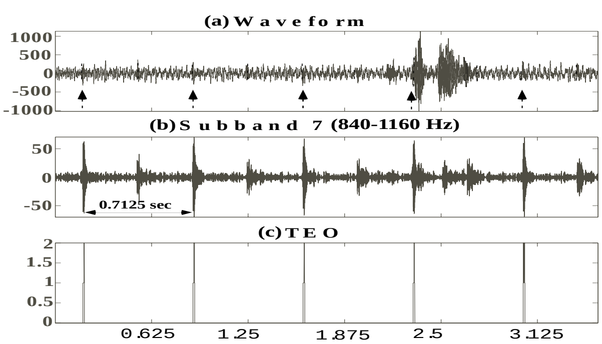


Fig. 7. Applying TEO processing to the environmental condition where the turn signal is on and the long-term noise is city driving with traffic, windows closed (N3).

5. DISCUSSION

The main goal of Environmental Sniffing is to extract knowledge about environmental noise that exists within continuous speech. As a first step towards this goal, in our evaluations, we focused on extracting knowledge about the acoustic environmental noise using a noise-only audio database. However, while constructing the framework, we provide sufficient flexibility to easily move towards a subsequent step and to allow the same framework to be used for noisy-speech sections in audio streams as well. We are presently working on a broad class monophone recognition based framework to extract environmental noise knowledge from an audio stream consisting of both noisy-silence and noisy-speech (e.g., similar to our speech activity detection [SAD] work previously reported [7]). After defining a set of broad phone classes (e.g. nasals, unvoiced fricatives, voiced fricatives, etc.), we can generate monophone model sets where each corresponds to a noise type by degrading the clean monophone models with noise. In addition to these models, a silence model will also be included for each noise type. If we use 10 broad class monophones, we will have 10 clean monophone models, 10xN noisy monophone models, 1 clean silence model and N noisy-silence models, for a total of $(N+1)$x11 models. Due to the pruning method used in the existing framework, the increase in search space will be less than a linear increase when we have more noise types. It will also be straightforward to use language modeling to calculate the transition probabilities from one monophone model set to another.

Another important issue is handling new in-coming noises within the framework, in other words, adapting Environmental Sniffing to new environmental noise types. Since there is a garbage noise model (N0) within the existing framework, we can keep track of the data classified as N0 and cluster to check if there is a sufficient data cluster to train a new additional noise model. We can also use the previous classification results to check how much the new model differs from existing ones by comparing the score distribution of the new model with existing ones.

6. CONCLUSION

In this paper, we have addressed the problem of changing acoustic environmental conditions in speech tasks. We proposed a new framework entitled *Environmental Sniffing* to detect, classify and, track changing acoustic environmental conditions and extract knowledge about the environmental noise. After proposing a general framework, we specialized the sniffer to an in-vehicle speech application. Novel aspects included a number of knowledge based processing steps such as T^2-BIC segmentation, noise language modeling, GMM classification and TEO processing. We believe such processing will provide significant knowledge to subsequent speech processing tasks and thereby increase robust speech performance.

7. REFERENCES

[1] M. A. Siegler, U. Jain, B. Raj, and R. M. Stern, "Automatic Segmentation, Classification and Clustering of Broadcast News Audio," *Speech Recognition Workshop (DARPA)*, Feb. 1997.

[2] R. Bakis, S. Chen, et al., "Transcription of BN Shows with the IBM LVCSR System," *Speech Recognition Workshop (DARPA)*, Feb. 1997.

[3] J.H.L. Hansen, P. Angkititrakul, J. Plucienkowski, et al., "CU-Move: Analysis & Corpus Development for Interactive In-Vehicle Speech Systems," *Eurospeech-2001*, vol. 3, pp. 2023-2026, Sept. 2001.

[4] B. Zhou, J.H.L. Hansen, "Unsupervised Audio Stream Segmentation and Clustering via the Bayesian Information Criterion," *Proc. ICSLP-2000*, vol. 3, pp. 714-717, Oct. 2000.

[5] B. L. Pellom, "Sonic: The University of Colorado Continuous Speech Recognizer," *Techical Report #TR-CSLR-2001-01*, March, 2001

[6] G. Zhou, J.H.L. Hansen, and J.F. Kaiser, "Nonlinear Feature Based Classification of Speech under Stress," *IEEE Trans. on Speech & Audio Processing*, vol. 9, no. 2, pp. 201-216, March 2001.

[7] R. Sarikaya, J.H.L. Hansen, "Robust detection of Speech Activity in the Presence of Noise," *Inter. Conf. on Spoken Language Processing*, vol. 4, pp. 1455-1458, Dec. 1998.

U. Yapanel, X. Zhang, J.H.L. Hansen, "High Performance Digit Recognition in Real Car Environments", ICSLP-2002:Inter. Conf. on Spoken Language Processing, vol. 2, pp. 793-796, Denver, CO USA, Sept. 2002

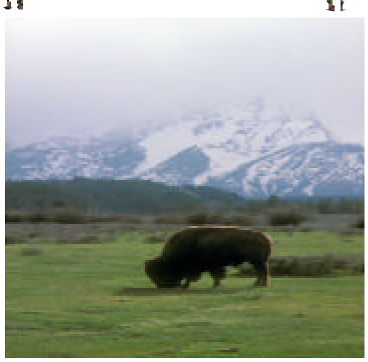
High Performance Digit Recognition in Real Car Environments

Umit Yapanel, Xianxian Zhang, John H.L. Hansen

CSLR RSPG

Robust Speech Processing Group
Center for Spoken Language Research
University of Colorado Boulder, Campus Box 594
(Express Mail: 3215 Marine Street, Room E-265)
Boulder, Colorado 80309-0594
303 – 735 –5148 (Phone) 303 – 735 – 5072 (Fax)
http://cslr.colorado.edu/
John.Hansen@colorado.edu, (email)

7th International Conference On Spoken Language Processing
ICSLP-2002
INTERSPEECH 2002



ICSLP-2002: Inter. Conf. on Spoken Language Processing
INTERSPEECH 2002
Denver, Colorado USA
September 16-20, 2002

High Performance Digit Recognition in Real Car Environments

Umit Yapanel, Xianxian Zhang and John H. L. Hansen

Robust Speech Processing Group, The Center for Spoken Language Research
University of Colorado at Boulder, Boulder, CO, 80309, U.S.A.
{yapanel,zhang,jhlh}@cslr.colorado.edu Web: http://cslr.colorado.edu

Abstract

In this paper, we consider the problem of robust digit recognition in real car environments. We choose to utilize newly-collected CU-Move database [2]. We address the problem using two integrated approaches . First, we consider array processing, enhancement and noise adaptation techniques as an integrated solution. This approach reduced the word error rate (WER) 38.6% and increased word accuracy (WAC) 47.1%, relative to baseline results. Secondly, we use array processing, enhancement, cepstral mean normalization, vocal tract length normalization and MLLR adaptation as an alternative solution. The net gain obtained with this solution is 55.4% reduction in WER and 64.3% increase in WAC, relative to baseline results. The first approach has the advantage of speed since all operations can be performed in real-time, while the second approach maintains high accuracy at the cost of increased computational requirements.

1. Introduction

The problem of robust speech recognition in car environments has attracted much attention, since command and control, number dialing and navigation through interactive system applications are of fundemantal importance. For hands-free cell phone use or in-vehicle car navigation, it is crucial to minimize driver's task stress so that appropriate cognitive function remains with operating the vehicle. Moreover, hands-free natural voice interaction with the vehicle offers the prospects of reduced distraction. These applications make high performance robust speech recognition in the car a necessity. However, speech recognition in car environments is fragile with word error rates (WERs) ranging from 20-60% depending on the road and vehicle conditions [2]. Several approaches to speech recognition in car includes combination of basic HMM recognizers with front-end noise supression, model adaptation to noise as well as speaker and multi-channel concepts. Many early approaches to speech recognition in car focused on isolated commands [4]. Other studies have shown improvement in computational requirements with front-end signal-subspace enhancement , with increases in recognition rates depending on the driving conditions [5], Another study [6] considered experiments on recognizer mismatch between training and testing using clean data and added car noise. This paper investigates the applicability and integrity of aforementioned techniques in a realistic environment, with the goal of still being close to processing in real-time.

The paper is organized as follows. In the next section, we describe the database followed by the baseline recognition system details. Section 4 considers the problem of array processing and describes two techniques. Section 5 is on front-end speech enhancement and its effect on recognition accuracy, the next section is devoted to noise adaptation techniques with special attention to Jacobian adaptation. Speaker adaptation techniques are considered next. After taking computational issues into account in section 8, we conclude the paper with a discussion and giving future research.

This work was supported by DARPA through SPAWAR under Grant No.N66001-8906.

2. CU-Move Database

The CU-Move project [3] aims to invent and develop car navigation systems that are reliable and employ a mixed-initiative dialog. This requires reliable speech recognition across changing acoustic conditions. There are 5 parts in the database; command and control words, digit strings being mostly phone numbers, street names with mostly spellings, phonetically balanced sentences and Wizard of Oz interactive navigation conversations. A total of 500 speakers produced over 600GB of data during the six month collection effort across the United States. The database and noise conditions are analyzed in detail in [2]. We point out that the noise conditions are changing with time and are quite different in terms of SNR, stationarity and spectral structure. The challenge in addressing these noise conditions is that they might be changing depending on the car being used and the road. In this study, we use the digits portion that contains approximately 40 words. In order to evaluate the noise level of each file, we utilize NIST's segmental SNR tool [12], which uses the audio file as well as voice activity detection file to determine an approximate noise level for the file. In Fig. 1, we present the variation of segmental SNR over time for one speaker from the digits portion (this portion of data collection includes recordings with windows open varying amounts). This figure reveals the time varying nature of speech interaction within the car and shows the difficulty of the problem. Even for a single speaker, segmental SNR can change by as much as 15 dB depending on the vehicle type, operating conditions (i.e. turn signal on, windows' positions, speech, wiper blades etc.) and road conditions. While this is a challenging task, it is somewhat matched since it is possible to collect speech for in-vehicle navigation dialogs across many possible road and driving conditions, even though the number of combinations can be quite large.The mean segmental SNR was approximately 10 dB for both train and test sets, again suggesting that the task is matched at least in terms of mean segmental SNR, though noise spectral contours can vary significantly.

3. Baseline recognition system

Although for a small vacabulary such as the 40-word digit task, it would be more appropriate to use a whole-word based system, we choose to use a sub-word based system to be able to

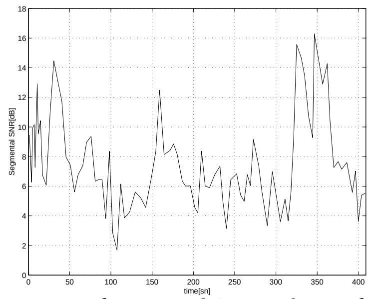


Figure 1: *Variation of segmental SNR with time for a speaker*

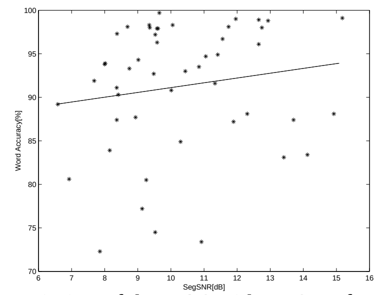


Figure 2: *Variation of the WAC with segSNR for the test set*

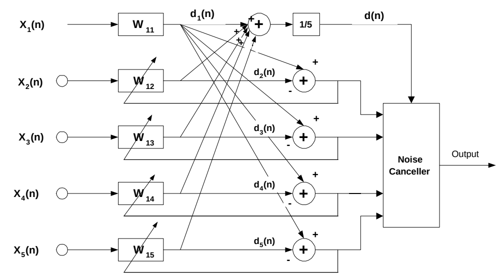


Figure 3: *Adaptive array processing*

generalize our results to tasks such as navigation dialogs using other parts of the database such as phonetically balanced sentences, street names or navigation dialog. Therefore, we use the Sonic continuous speech recognizer [1] being developed at CSLR for a number of recognition tasks. The recognizer is based on a cross-word continuous denstiy HMM using acoustic models that employ decision-tree state clustered HMMs with associated Gamma densities for duration distributions. The decoder implements a two-pass token passing approach to coninuous speech recognition [1].

A total of 60 speakers balanced across gender and age (18-70 yrs. old) were used in the training set. However, before training, it was first necessary to perform forced alignment on the entire training corpus using Sonic's alignment tool [1]. We checked several alignments visually by hand. Although the data was very noisy, the alignments were in perfect agreement with what a human trancriber would produce. The 39 dimensional feature set contains 12 MFCCs, deltas and delta deltas along with c0, delta and delta delta energy. The reason for extracting c0 in place of energy was to be able to use model adaptation algorithms that require a conversion back to the linear spectral domain. The test set also contained a balanced gender and age set of 50 speakers. The HMMs were trained using Sonic decision-tree HMM trainer resulting in 444 models. All HMMs have left-to-right topology with no skips and each state was represented by 6-24 mixtures depending on the available training data. The vocabulary size was 40 including silence (SIL) and unkown word (UNK). The dictionary is very convenient for telephone dialing applications since it contains most necessary words like "dash", "pound", "sign" in addition to numbers. Under these conditions, the baseline results for the recognizer is given in Table 1. The results are good if we consider the varying noise level

Table 1: *Baseline recognition results.*

Type	Score(%)	**Rel. Imp.(%)**
WAC	91.3	**0.0**
WER	10.1	**0.0**

of the speech. These variations are mostly due to varying window position and speeds occured during speech data collection. To see the correlation of the WAC with segmental SNR, we produced the scatter plot in Fig. 2, which shows some correlation with the SNR.We might, therefore, expect some improvement using noise adaptation techniques, but it seems that the performance might better improve if we are able to employ speaker adaptation techniques as well, since there is a wide range of WAC at similar segSNR levels.

4. Array Processing

In the formulation of the CU-Move Corpus, an array of five microphones was designed and constructed to allow array processing techniques to improve SNR and hence recognition accuracy. In this section, we consider two array techniques and give recognition results for delay-and-sum beamforming (DASB). In DASB, the position of the speaker is assumed to be fixed and for each microphone, a sample delay is computed. Each microphone signal is steered with the corresponding delay and summed to yield an enhanced signal. This approach has the drawback that as the head position of the speaker changes, the delay terms may not be optimal. Therefore, techniques that adaptively adjust the delay terms may yield better results. In Table 2, we present the recognition results with the front-end DASB scheme. The relative improvement in accuracy is 20.7%, which implies that better results could be obtained by improving the design of the array processing technique. Since every

Table 2: *Recognition results after DASB.*

Type	Score(%)	**Rel. Imp.(%)**
WAC	93.1	**20.7**
WER	8.2	**18.8**

driver moves his/her head during driving, an adaptive array processing (AAP) approach that detects the movements and adjusts the time delays accordingly and automatically is useful. Therefore, tracking speaker's head movement will increase the accuracy of the array processing. The proposed AAP algorithm is summarized in Fig. 3. In this algorithm, we choose microphone 1 as the primary microphone, and build an adaptive filter between it and each of the other four microphones. Next, we sum and scale all the steered microphone outputs to provide an enhanced signal (d[n]). An additional processing feature for the adaptive algorithm is the noise canceller shown in Fig. 3. We again sum the outputs in adjacent pairs to provide three noise reference signals. These signals are used as inputs to a three-channel adaptive noise canceller, the output of which is substracted from the output of AAP to yield the enhanced output. To evaluate the AAP system, we selected ten speakers from the CU-Move database that were balanced across gender and age. Each speaker was processed using the delay-and-sum beamformer (DASB), adaptive array processor (AAP), and adaptive array processor plus noise canceller (AAP+NC). The segSNR tool [12] was again used to evaluate noise suppression performance. Table 3 shows average improvements compared to unprocessed center channel 3 for ten speakers.

Table 3: *SegSNR improvements of array processing methods.*

Method	DASB	AAP	AAP+NC
Ave. imp. [dB]	-0.045	1.01	2.05

5. Front-End Speech Enhancement

Since the recognition problem is related to time-varying noise, applying a fast enhancement technique before recognition may be beneficial in two ways. First, it can reduce the mismatch between train and test data. Second, and more importantly, it can reduce the variance between different noise conditions and help increase the efficiency of the noise adaptation algorithms to be discussed in the next section. As a fast and effective enhancement algorithm, we choose to use MMSE [9]. A listener evaluation of the enhanced files was conducted to assess the effect of the enhancement. Subjective results showed the enhancement to be very effective in terms of reducing the perception of noise. The front-end enhancement method was used for a recognition test of CU-Move data, with noise estimated from the silence portion of the files (generally the first 300 msec.). The results presented in Table 4 shows that the front-end enhancement is also effective in reducing the WER by 25.8% relative to the baseline recognition system.

Table 4: *MMSE enhancement results after DASB*

Type	Score(%)	**Rel. Imp.(%)**
WAC	94.2	**33.3**
WER	7.5	**25.8**

6. Noise Adaptation

Another approach to improve recognition performance in noisy conditions is to use noise adaptation techniques [10, 8]. Among many approaches proposed, we find Jacobain Adaptation (JA) particularly useful and suitable for this application. Parallel Model Combination(PMC) requires initial clean models and then, using a portion of noise only data, successfully adapts the models to the noisy conditions. However, in our case, we do not have access to clean models. JA, on the other hand, assumes noisy models trained in one condition and adapts models to a target condition using a sample of the target noise data. Moreover, adaptation is very fast; therefore, JA is suggested as a promising approach to the problem of robust speech recognition in the car.

6.1. Jacobian Adaptation

Jacobian adaptation assumes that acoustic models are nonlinear functions of speech and additive noise, then a change in the noise conditions will affect the models in a nonlinear fashion. We assume that we have two noise conditions, namely a reference noise in which the current acoustic models were trained, and, target noise contained within the incoming contaminated speech. An important condition is that the spectral change in the noise condition should be small so that the noisy speech model statistics stay within the linear range of the Jacobain adaptation. We can briefly explain Jacobian adaptation as follows. Assume we have a vector Y which is an analytic function of X. A small change ΔX in X causes a small change ΔY in Y. This change can be expressed as in Eqn. (1).

$$\Delta Y = \frac{\delta Y}{\delta X}\Delta X \tag{1}$$

A similar equation can be written for the cepstrum of the noisy speech in the reference noise as Eqn. (2),

$$C_{S+N_r} = \frac{\delta C_{S+N_r}}{\delta C_{N_r}}\Delta C_{N_r} \tag{2}$$

The term with partial derivatives is the Jacobian that can be computed in advance and stored to be used during model adaptation. The Jacobian can be expressed in terms of linear spectra as in Eqn. (3),

$$C_{S+N_r} = F^* \frac{N_r}{S+N_r}F, \tag{3}$$

where F is the inverse discrete cosine transform matrix, N_r is the reference noise spectrum and $S+N_r$ is the noisy speech spectrum contaminated by the reference noise. Three operations are required during the adaptation;(1) collecting noise statistics, (2)multiplying the difference between the reference and target noise statistics to obtain the bias term, and (3) adding this bias term to the noisy cepstrum (in the reference noise) to obtain the noisy speech cepstrum in the target noise as in Eqn. (4),

$$C_{S+N_t} = C_{S+N_r} + \frac{\delta C_{S+N_r}}{\delta C_{N_r}}(C_{N_t} - C_{N_r}). \tag{4}$$

Therefore, we only need a matrix-vector multiplication to compute the bias and add this bias to the existing acoustic models to obtain the adapted models. This limited computational requirement makes this technique viable for real-time applications. However, applicibility of the JA is also limited for this task. Because it requires training in one noise condition, which does not hold in our case since the training data contains many different noise types and levels. This causes degradation in the performance but still gives a good amount of improvement in comparison with no adaptation case. An ideal application would be to train acoustic models in *idle* noise conditions and then test against many different noise conditions. This will yield much better results since the models are not as wide in comparison with our case which includes a variety of noise types and levels.

Table 5: *JA results after DASB and MMSE enhancement*

Type	Score(%)	**Rel. Imp.(%)**
WAC	95.4	**47.1**
WER	6.2	**38.6**

7. Speaker Adaptation

At this point, we have reached a level of performance by applying three distinct techniques, in a cascade fashion, namely array processing, enhancement and noise adaptation. However, as we saw in Fig. 2, the problem is also related to speaker variability. To explore what impact speaker adaptation can have for this problem, we choose to use two different processing techniques, namely vocal tract length normalization (VTLN) and Maximum Likelihood Linear Regression (MLLR). Since these techniques do not require a conversion back to the linear spectrum, we first apply cepstral mean normalization (CMN) to limit any channel effects that may stem from data collection as well as from array processing.

7.1. Cepstral Mean Normalization

CMN is a well-known technique that is especially efficient for removing long term channel effects. We normalized 12 MFCCs with a mean over a 300-frame buffer(i.e. 360 ms) and then compute the delta and delta delta terms [1]. The results after DASB and enhacement are summarized in Table 6. The absolute gain,in comparison with no CMN case, is 1% in accuracy and 1.2% in WER.

Table 6: *CMN results after DASB and MMSE enhancement*

Type	Score(%)	**Rel. Imp.(%)**
WAC	95.2	**44.8**
WER	6.3	**37.6**

7.2. Vocal Tract Length Normalization

Another good approach to reducing speaker variability is vocal tract length normalization(VTLN) [11]. VLTN was used in a cascade manner with CMN and improved results were obtained as summarized in Table 7. It is believed that the results would be better if we were able to more accurately determine the vocal tract length of each speaker, however, inaccuricies in determining this length reduces the overall performance. In spite of this, the level of improvement is still considerable.

Table 7: *VTLN results after DASB, MMSEE and CMN*

Type	Score(%)	**Rel. Imp.(%)**
WAC	96.0	**54.0**
WER	5.4	**46.5**

7.3. Maximum Likelihood Linear Regression

MLLR [13] was originally proposed for speaker adaptation, however, if the speech data is noisy then it also adapts to noise, and is therefore a good candidate for combining of noise and speaker adaptation. Our previous simulations with the DARPA SPINE I&II tasks [7] also suggested that MLLR is an integrated way of adapting the models to both speaker and noise variations. The results after 3 iterations of MLLR are presented in Table 8.

Table 8: *MLLR results after DASB, MMSE and CMN*

Type	Score(%)	**Rel. Imp.(%)**
WAC	96.9	**64.3**
WER	4.5	**55.4**

The improvement is remarkable. However, there are some problems with the computational requirements as we discuss in the next section.

8. Computational Issues

Our ultimate goal is to be able to perform speech recognition in car environments in real-time. Therefore, in this section, we summarize computation times required for various methods utilized. In all of the simulations, we used a single processor machine with 512 MB of RAM and a 1.5 GHz processor speed. Table 8 summarizes these computation times. A good combination in terms of computation time and accuracy is DASB+MMSE enhancement+JA requiring a real time factor of 0.5. This means that we still have half of the time to perform additional processing. This time can be devoted to more effective array processing techniques and making the model adaptation more accurate and efficient.

Table 9: *Computation times*

Operation	**Real-time factor**
Recognition	0.23
DASB	0.10
MMSE enhancement	0.10
JA	0.06
CMN	0.02
VTLN	0.58
MLLR	1.83

9. Discussion and Research Directions

In this paper, we presented our results on the newly introduced CU-Move corpus. We applied a variety of techniques to the problem of robust speech recognition in the car in a cascade fashion to obtain improved results. Another concern was to be able to achieve speech recognition in real-time. We showed that with fast adaptation techniques, real-time processing is viable. As for future research directions, we are working on more involved array processing techniques to obtain as much noise suppression as possible. Another improvement could come from incorporating fast speaker adaptation techniques such as SMLEM [14] into recognition together with noise adaptation. Lastly, we are working on moving to other in-vehicle tasks requiring larger vocabularies than the digits portion to see the applicability and viability of large vocabulary continuous speech recognition in real car environments.

10. References

[1] B. Pellom, ”Sonic: The University of Colorado Continuous Speech Recognizer”, *Tech. Rep.TR-CSLR-2001-01*, CSLR, Univ. of Colo., March 2001.

[2] J.H.L. Hansen, P. Angkititrakul, J. Plucienkowski, S. Gallant, U. Yapanel, B. Pellom, W. Ward, R. Cole, “CU-Move: Analysis & Corpus Development for Interactive In-vehicle Speech Systems”, *Eurospeech’01*, Sept. 2001

[3] http://cumove.colorado.edu

[4] S. Riis, O. Viikki,”Low-complexity Speaker Independent Command Word Recognition in Car Environments”, *IEEE ICASSP’00*, 1743-46

[5] J. Huang, Y. Zhao, S. Levinson,”A DCT-based Fast Enhancement Technique for Robust Speech Recognition in Automobile Usage”, *Eurospeech’99*,1947-1950

[6] A. Fischer, V. Stahl,”Database and Online Adaptation for Improved Speech Recognition in Car Environments”, *IEEE ICASSP’99*

[7] J.H.L. Hansen, R. Sarikaya, U. Yapanel, B. Pellom, ”Robust Speech Recognition in Noise: An Evaluation using the SPINE Corpus”, *Eurospeech’01*, Sept.2001

[8] R. Sarikaya, J.H.L. Hansen, “PCA-PMC: A Novel Use of a Priori Knowledge for Fast Parallel Model Combination” *ICASSP’00*, June 2000

[9] Y. Ephraim, D. Malah,”Speech Enhancement Using a Minimum Mean Square Error Short Time Spectral Amplitude Estimator”, *IEEE Trans. ASSP-32*, 1109-21, 1984

[10] S. Sagayama, Y. Yamaguchi, S. Takahashi and J. Takahashi, “Jacobian Approach to Fast Acoustic Model Adaptation”, *ICASSP’97*, 835-8, 1997

[11] L.F. Uebel, P.C. Woodland, “An Investigation into Vocal Tract Length Normalization”, *Eurospeech’99*, 1999

[12] SPHERE software package, www.nist.gov

[13] C.J.Legetter, P.C.Woodland,”Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models”, *Computer Speech & Language*, Vol.9, 171-85, 1995

[14] B. Zhou, J.H.L. Hansen,”A Novel Algprithm for Rapid Speaker Adaptation Based on Structural Maximum Likelihood Eigenspace Mapping”, *Eurospeech’01*, 1215-8