

## SPEECH, EGG, AND GLOTTAL AREA DATA

The data described herein are contained on two CD-ROMs. One CD-ROM contains the Normal folders, while the other CD-ROM contains the Disorder folder, the Mimic folder, the Other folder, the Additional folder, and the Area folder.

### A4.1 Normal FOLDERS. PROCEDURES AND DATA

All data recordings were performed in a professional single-wall sound room. The speech and electroglottographic (EGG) signals were monitored synchronously and simultaneously. One of two microphones was used: an Electro-Voice (Buchanan, Michigan) RE-10 dynamic cardioid or a Bruel and Kjaer (A Division of Spectris Technologies, Inc., Norcross, GA) model 4113 condenser. The selected microphone was located 6 in. from the speaker's lips. The electroglottograph was a Synchronvoice, Inc. model. All data were directly digitized, thereby avoiding any low-frequency distortions that may have been introduced through the use of audio tape recordings. The speech and EGG signals were bandlimited to 5 kHz by antialiasing elliptic filters with a minimum stop-band attenuation of  $-55$  dB and a passband ripple of  $\pm 0.2$  dB. Both signals were amplified by a Digital Sound Corp. (Maynard, Massachusetts) DSC-240 audio control console. The two signals were sampled at 10 kHz per channel by a Digital Sound Corp. DSC-200 analog-to-digital system with 16-bit resolution. The data that were recorded using the Electro-Voice microphone were corrected for microphone distortions by deriving a microphone correction transfer function (Childers and Wong, 1994). The data that were recorded with the Bruel and Kjaer microphone did not require correction since its bandwidth characteristics were sufficiently broad enough that no frequency distortions were introduced into the data.

The total number of subjects was 52 (25 male, 27 female) with normal larynges. The subject's ages ranged from 20 to 80 years old. The complete speech protocol consisted of 28 tasks, including twelve sustained vowels /Y, IH, EY, EH, AE, UW, UH, OW, AO, AA, AH, ER/; five sustained unvoiced fricatives /HH, F, TH, S, SH/; and four sustained voiced fricatives /V, DH, Z, ZH/. This notation is the upper case version of the ARPAbet. The subjects were instructed to pronounce and sustain each vowel as it would be pronounced in the following words, respectively: bet, bit, bait, bet, bat, boot, book, boat, bought, Bach, but, Bert. For the fricatives, we used the following cue words: hat, fix, thick, sat, ship, van, this, zoo, and azure. The duration of each vowel and fricative approximated 2 sec. The additional tasks include counting from one to ten with a comfortable pitch and loudness, counting from one to five with a progressive increase in loudness, singing the musical scale using "la," and speaking three sentences (We were away a year ago; Early one morning a man and a woman ambled along a one mile lane; Should we chase those cowboys?). All of the above 27 tasks were recorded using the Electro-Voice RE-10 microphone. The 28th task used the Bruel and Kjaer model 4113 microphone, and repeated the sustained vowel, /AA/, as in Bach.

The data from subjects with a normal larynx are contained on the CD-ROM in the folders denoted as follows:

Normal\_m folder (folder containing data folders for males) m01 to m25 folders

Normal\_f folder (folder containing data folders for females) f01 to f27 folders

Each male (female) folder contains 56 data files: 28 speech and 28 EGG. The files are labeled mxxyyz.dat for male (with a similar notation for female), where xx denotes the subject number, yy denotes the speech task, and z denotes whether the file is speech or EGG data.

xx: ranges from 01 to 25  
 yy: ranges from 01 to 28  
 z: is either s (speech) or e (EGG)  
 dat: is the extension, denoting ASCII format

yy denotes the tasks, which were as follows.

- 01: counting from 1 to 10 with a comfortable pitch and loudness
- 02: counting from 1 to 5 with a progressive increase in loudness
- 03: sustained vowel /IY/ as in beet
- 04: sustained vowel /IH/ as in bit
- 05: sustained vowel /EY/ as in bait
- 06: sustained vowel /EH/ as in bet
- 07: sustained vowel /AE/ as in bat
- 08: sustained vowel /UW/ as in boot
- 09: sustained vowel /UH/ as in book
- 10: sustained vowel /OW/ as in boat
- 11: sustained vowel /AO/ as in bought
- 12: sustained vowel /AA/ as in Bach
- 13: sustained vowel /AH/ as in but
- 14: sustained vowel /ER/ as in Bert
- 15: sustained fricative /HH/ as in hat
- 16: sustained fricative /F/ as in fix
- 17: sustained fricative /TH/ as in thick
- 18: sustained fricative /S/ as in sat
- 19: sustained fricative /SH/ as in ship
- 20: sustained fricative /V/ as in van
- 21: sustained fricative /DH/ as in this
- 22: sustained fricative /Z/ as in zoo
- 23: sustained fricative /ZH/ as in azure
- 24: produce the musical scale singing "la," attempting to go up then down as one effort with a pause between the top two notes

Orally read the following sentences with a comfortable pitch and loudness.

- 25: We were away a year ago.
- 26: Early one morning a man and a woman ambled along a one mile lane.
- 27: Should we chase those cowboys?
- 28: Sustained vowel /AA/ as in Bach using a Bruel and Kjaer microphone, model 4113 condenser.

Thus, the example data file m0110s.dat would be the sustained vowel /OW/, as in boat, while m0110e.dat would be the EGG signal for the same sustained vowel.

The speech and EGG data were recorded simultaneously. However, the speech data lags the EGG data by the amount of time it takes for the sound to propagate from the vocal folds to the microphone (6 inches beyond the lips). This propagation delay varies with each subject due to variations in vocal tract length. A "typical" value in data samples would be 7 data samples at 10 kHz sampling frequency. This approximates the distance of 24 cm from the vocal folds to the microphone. Consequently, to align the speech and EGG data on a sample basis, shift the speech data to the left by 7 data samples while keeping the EGG data fixed (or shift the EGG data to the right by 7 data samples).

## Missing Data

Several files are missing: f0801s.dat, f0802s.dat, f0803s.dat, f0828e.dat, f0828s.dat, f0928e.dat, f0928s.dat, f1028e.dat, and f1028s.dat.

## A4.2 Disorder FOLDER PROCEDURES AND DATA

---

The data for the subjects with vocal disorders were collected in a manner identical to that described for the subjects with a normal voice. The file naming convention follows that specified for the normal data set. However, note that the subjects are not the same and the tasks were slightly different. I, however,

- 01: Counting from 1 to 10 with a comfortable pitch and loudness
- 02: Counting from 1 to 5 with a progressive increase in loudness
- 03: Sustained vowel /IY/ as in beet. (Some subjects phonated three vowels in succession: /IY/, /AA/, /UW/. This can be determined by examining the speech file.)
- 04: Produce the musical scale singing "la," attempting to go up then down as one effort with a pause between the top two notes

Orally read the following sentences with a comfortable pitch and loudness.

- 05: We were away a year ago.
- 06: Early one morning a man and a woman ambled along a one mile lane.
- 07: Should we chase those cowboys?

The vocal disorder database consists of 16 patients with vocal disorders. Two patients were remeasured twice and 3 patients were remeasured once after treatment. A description of the vocal disorders can be found in Table A4.1 below. The range of voices varied from mildly deviant to very deviant as determined by a clinician. The patients were asked to phonate the vowel /IY/ for about 2 sec.

The file names follow the convention adopted for the Normal data set, namely mxyyz.dat for male speakers (with a similar notation for female speakers).

- xx: ranges from 01 to 08 for male and from 01 to 15 for female
- yy: ranges from 01 to 07
- z: is either s(speech) or e (EGG)
- dat: is the extension denoting ASCII data

## A4.3 Mimic FOLDER

---

In addition to the above disorder data, this folder contains a set of data that mimics vocal disorders.

- m01xxms.dat: Male subject 01, task xx, m (mimic), s (speech). The tasks range from 01 to 07. The data in this set is a normal modal voice.
- m02xxms.dat: This is the same subject as above but the subject mimicked a breathy voice for all tasks.
- m03xxms.dat: Male subject 03, task xx, m (mimic), s (speech). The subject mimicked a vocal fry for all tasks.
- m04xxms.dat: This is the same subject as m03. The subject mimicked a breathy voice for all tasks.
- m05xxms.dat: Male subject 05, task xx, m (mimic), s (speech). The subject mimicked a hoarse voice for all tasks.
- m06xxms.dat: Same subject as m05. The subject mimicked "computer voice" for all tasks.

**TABLE A4.1. List of Vocal Disorder Speakers**

Subject number	Symptoms
m01	True vocal cords (TVC) contact ulcer
m02	Hoarse
m03	Vocal fry
m04	Breathy; hoarse
m05 <sup>a</sup>	Hoarse unilateral TVC carcinoma
m06	m04 1 month post-injection
m07	m02 5 months post-injection
m08 <sup>a</sup>	Bilateral paralysis of TVC
f01 <sup>a</sup>	Posterior cyst
f02	Bilateral nodule
f03	Mild hoarseness
f04 <sup>a</sup>	Unilateral paralysis
f05	f04 1 month post-injection
f06 <sup>a</sup>	Hyper functional
f07	Breathy, weak
f08 <sup>a</sup>	Vocal fry
f09 <sup>a</sup>	f04 4 months post-injection
f10	f08 1 month post-injection
f11	f07 1 month post-injection
f12 <sup>a</sup>	Enlarged vocalus muscle
f13 <sup>a</sup>	Right TVC unilateral paralysis
f14	f08 3 months post-injection
f15	Nodules

<sup>a</sup>Very deviant voices.

Missing files: f0901e and f0901s, m0707e and m0707s.

m07xxms.dat: Same subject as m05. The subject mimicked a rough voice for all tasks.

m08xxms.dat: Same subject as m01. The subject mimicked a rough voice for all tasks.

Note: subjects m01, m02, and m08 are the same male. Subjects m03, m04, m05, m06, and m07 are the same male, but differs from m01. These two mimics are professional voice clinicians with many years experience and are recognized as expert mimics of various vocal disorders. Missing files include m0602me, m0602ms, m0604me, and m0604ms.

Note: the EGG data are denoted with an e in place of the s.

## A4.4 Other FOLDER

The data contained in this folder contains only speech data for six adult speakers (three male and three female) each speaking the following sentence.

That zany van is azure. (voiced fricatives)

The data were collected as described for the Normal data set. Note that this data set does not include the EGG signal. The subjects had no vocal disorders.

The data files are denoted as follows.

m01zs.dat: Male subject 01, sentence: That zany van is azure, (s) speech.

m01ze.dat: Male subject 01, sentence: That zany van is azure, (e) EGG.

f01zs.dat: Female subject 01, sentence: That zany van is azure, (s) speech.

These files continue to be denoted but are not shown here. The subjects for this data correspond to the subjects listed for the Normal data as follows.

Other folder	Normal folder
m01	m01
m02	m03
m03	m05
f01	f06
f02	not in Normal set
f03	f01

#### **A4.5 Additional FOLDER**

---

Nine sentences were recorded. The same male speaker spoke each sentence twice. The male speaker is not in the Normal set. The file naming convention is such that the first number denotes the sentence number, which is followed by an underscore and a second number that denotes the repetition. Thus, there are eighteen speech and corresponding EGG data files in this corpus. The file name starts with "spe" if it is a speech file; it starts with "egg" if it is an EGG file.

An example data file name for sentence 1 is the following.

egg1\_1.dat for the egg file for sentence 1, first repetition of two  
 egg1\_2.dat for the egg file for sentence 1, second repetition of two  
 spe1\_1.dat for the speech file for sentence 1, first repetition of two  
 spe1\_2.dat for the speech file for sentence 1, second repetition of two

1. The rain in Spain falls mainly on the plain.
2. A bird in the hand is worth two in the bush.
3. The supplies were stored for everyone.
4. Early one morning a man and a woman ambled along a one mile lane.
5. We were away a year ago.
6. I owe you a yo-yo.
7. Should we chase those cowboys?
8. The zany van is azure.
9. We saw the ten pink fish.

Sampling frequency: 10,000 Hz

Data format: 16-bit ASCII.

#### **A4.6 Area FOLDER**

---

The data in this folder contains the glottal area measured from ultra-high speed laryngeal films, the EGG signal, and the speech signal. These data were measured simultaneously. The procedures are described in Childers and Krishnamurthy (1985). Also see the references in this paper for additional detail.

#### **A4.7 PROCEDURES AND AREA DATA/Normal**

---

The subjects were four adult males, coded as m01, m02, m03, and m04, who had no vocal disorders. The experimental tasks for each subject consisted of phonation of the vowel /IY/ at three different intensities at each of three different fundamental frequencies. The recorded phonation was sustained for about three seconds. The three target fundamental frequencies used were 125 Hz, 170 Hz, and

340 Hz. To control the fundamental frequencies during the experiments, the subjects were asked to match a pure tone of the appropriate frequency that they heard over a pair of headphones. The three different intensities at each fundamental frequency represent a "comfortable" intensity, an intensity approximately 4 dB above the comfortable level, and another intensity about 4 dB below the comfortable level. The actual intensities were monitored with a sound meter. Thus, there were nine tasks for each subject, giving a total of 36 tasks for the four subjects.

The procedures are described in Childers and Krishnamurthy (1985) and references. Briefly, each subject was filmed using ultra-high speed photography. The EGG waveform was filmed simultaneously as a trace from an oscilloscope. This trace appeared on the film along with a timing signal. The EGG signal and the speech signal were also tape recorded along with the same timing signal. The microphone was attached to the handle of the laryngeal mirror at the point where the mirror frame joins the handle. The distance of the glottis from this point varies from subject to subject, but was approximately 11 cm. The audio bandwidth of the microphone was about 6 kHz.

For these data, the glottal area and traced EGG have been aligned as a pair of signals. Similarly, the tape recorded EGG data and speech data have been aligned as a pair of signals. However, these two data set pairs have not been aligned, but can be if desired. Perhaps the simplest procedure to align the two pairs of signals is to shift one pair relative to the other until the EGG waveforms are aligned.

The glottal area waveform is sampled at 5 kHz, so it must be interpolated to twice its data record length before it can be compared (plotted) with the other three data records. After interpolation, the glottal area waveform will be approximately 250 samples, which is approximately the same length as the traced EGG waveform. These two waveforms are much shorter in duration than the taped EGG and speech data, which are several seconds in duration.

All four waveforms can be plotted as a superimposed figure or as four separate figures for comparison purposes.

The convention for the file names for subject m01 is as follows.

am01 nn.dat  
tm01 nn.dat  
em01 nn.dat  
sm01 nn.dat

The first letter denotes the following.

a: glottal area data sampled at 5 kHz  
t: traced EGG data sampled at 10 kHz  
e: EGG data sampled from tape at 10 kHz  
s: speech data sampled from tape at 10 kHz

The letters nn denote the tasks which totaled nine.

The file names less the leading preface (a, t, e, s) are:

Filename	Measured Intensity (dB)	Target Frequency (Hz)	Measured Frequency (Hz)
m0101.dat	64	125	186.6
m0102.dat	68	125	198.0
m0103.dat	72	125	189.0
m0104.dat	68	170	156.3
m0105.dat	70	170	214.0
m0106.dat	80	170	
m0107.dat	70	340	201.0
m0108.dat	74	340	202.0
m0109.dat	79	340	330.9
m0201.dat	68	125	(The e file is missing.)
m0202.dat	74	125	(The e file is missing.)

(Continued)

<i>Filename</i>	<i>Measured Intensity (dB)</i>	<i>Target Frequency (Hz)</i>	<i>Measured Frequency (Hz)</i>
m0203.dat	77	125	114.5
m0204.dat	64	170	162.6
m0205.dat	70	170	161.3
m0206.dat	74	170	162.6
m0207.dat	63	340	333.3
m0208.dat	68	340	356.0
m0209.dat	74	340	336.1
m0301.dat	52	125	
m0302.dat	56	125	161.3
m0303.dat	60	125	
m0304.dat	64	170	161.3
m0305.dat	70	170	175.4
m0306.dat	76	170	180.2
m0307.dat		All files missing for this task.	
m0308.dat		All files missing for this task.	
m0309.dat	73	340	321.1
m0401.dat	56	125	122.9
m0402.dat	60	125	
m0403.dat	64	125	
m0404.dat	68	170	
m0405.dat	72	170	165.3
m0406.dat	75	170	
m0407.dat	54	340	277.7
m0408.dat	58	340	330.6
m0409.dat	62	340	346.0

## A4.8 PROCEDURES AND AREA DATA/Disorder

Also contained in this folder are data for three patients with vocal disorders. The procedures are the same as that for the normal data. The subjects were three adults, one female and two males. There was only one task for these subjects, namely to phonate the vowel /IY/ at a comfortable intensity and fundamental frequency. The same data were collected as for the normal subjects. The file names are similar to those used for the normal subjects: a prefix of a, t, e, or s; the subject's number (f01, f05, f06, f06); task number (01 for all subjects); and the extension (.dat). Thus, there are four files for each subject, for a total of 12 files. The vocal disorders for these patients are as follows.

- f01: Unilateral nodule, which has an observable affect on the EGG waveform.
- m05: Unilateral polyp with companion bulge on the other fold, which results in some loss of voice.
- m06: Voice production problem, loss of voice.

## A4.9 REFERENCES

Childers, D. G., and Krishnamurthy, A. K. (1985) A critical review of electroglottography. *CRC Critical Reviews in Bioengineering*, **12**, 131-161.

Childers, D. G., and Wong, C. F. (1994) Measuring and modeling vocal source-tract interaction. *IEEE Transactions on Biomedical Engineering*, **41**, 663-671.