

Development of a cost effective configuration for optimal quality recording for speech analysis

by Ajeesh Abraham

FILE	DR_AJISH.DOC (81K)	WORD COUNT	2115
TIME SUBMITTED	21-JUL-2016 01:39PM	CHARACTER COUNT	11325
SUBMISSION ID	690871030		

3
PROJECT PROPOSAL FORMAT

Part - A

1.0 Title of the project : **Development of a cost effective configuration for optimal quality recording for speech analysis.**

Area of Research : e) Electronics & Acoustics

2
1.1 Principal Investigator : Dr. Ajish K Abraham

1.2 Principal Co-Investigator : Dr. Vasantha Lakshmi M S

2
1.3 Collaborating Institution : Nil

1.4 Total Grants required (in figures and in words) : Rs. 5,00,000/-
(Rupees Five lakhs only)

1.5 Duration of the project : Twelve months

2.0 Project summary

Recording of speech samples is the first and important step in assessment, diagnosis, monitoring of progress in rehabilitation and research in the field of communication disorders. A reliable, user friendly, widely available and cost effective recording protocol is essential for achieving the desired objectives of all these activities. An evaluation of the reliability across recording methods requires consideration of hardware, protocols and acoustic analysis to determine the closeness of recorded speech sample to original one. **6** Factors affecting the quality of recording were investigated in the past by several researchers. The outcome of one of these suggests that, **6** highest quality recording would be made using a standalone digital recorder with high quality microphone in a sound treated room. This setup is quite expensive, unaffordable and not readily available for most of the researchers and clinicians. Hence, there is a need to find an alternative widely available and cost effective configuration with optimal quality recording. This project aims to develop and validate a cost effective combination of microphone, digitizing system (sound card), processing system (laptop) etc. which will result in optimal quality recording. The effective combination would be arrived at through investigation of the influence of each of the hardware components and associated factors by comparing the spectral, temporal and spectro-temporal parameters of the recorded

speech with the speech recorded by a bench mark system. The bench mark system for comparison would be a precision recording system connected with a higher end condenser microphone. Spectral parameters to be considered in the study include the fundamental frequency, first, second and third formants. Temporal parameters include Voice Onset Time (VOT), Closure Duration (CD) and Transition Duration (TD). Spectral tilt is the spectro-temporal parameter considered for comparison. Voiced and unvoiced stop consonants paired with vowel /a/ will be used as the speech stimuli. Ten participants (five males and five females) in the age range of 21-36 will be instructed to utter the stimuli with an appropriate loudness in a natural manner. The uttered stimuli will be recorded by the bench mark system and all the configurations considered for the study. The optimal configuration identified in comparison with the bench mark system would be validated by another set of six participants. Outcome of the study will be a validated protocol for recording speech samples with optimum quality.

3.0 Introduction

3.1 Definition of the problem

Many researchers have identified and compared the factors affecting the quality of sound recording, but none of these researchers have arrived at an optimal combination of hardware components and processing features which would result in a quality recording. Hence, there is a need to identify and develop a cost effective configuration for optimal recording of speech samples.

3.2 Objectives

- a. Investigate the effect of different laptop configurations, sound cards and microphones on the fundamental frequencies, formants, durations and spectral tilt of the speech samples recorded from the participants in comparison with that of the same samples recorded by a bench mark system.
- b. Compare and identify the most significant factor/factors individually and in combination.
- c. Develop a cost effective configuration for optimal quality recording.

- d. Validation of the developed configuration on the recording of different group of subjects.

2

3.3 Review of status of research and development in the subject

It has been always difficult for a researcher or a clinician to arrive at an optimal recording configuration while recording speech samples for acoustic analysis, for the purpose of assessment and diagnosis. Factors affecting the quality of such recordings were investigated in the past. Parsa, Jamieson & Pretty (2001) probed the effect of the type of microphone on acoustic measures and found that, each of the microphone was modifying the frequency spectrum of the voice signal. Livijn (2004) investigated the acoustic effects of recording speech samples through four different media and the implications for forensic phonetics. He found that, the cassette tape recorder was effectively preserving formant frequencies compared to other media. Deliyski, Shaw & Evans (2005) investigated the influence of sampling rate on accuracy and reliability of acoustic voice analysis and recommended a sampling rate of 26 kHz and above. Vogel and Morgan (2009) investigated the influence of microphone, environmental noise, Analog to Digital conversion and file format on the quality of recording and found that, all these factors have significant influence. Vogel, Rosen & Reilly (2014) compared four recording devices. Speech samples were recorded simultaneously using a hard disk recorder coupled with a mixer and a table mounted speaker, a landline telephone, smart phone, and a laptop computer with USB connected mic. They found that, the parameters of each of these recordings were significantly different from those of original speech samples.

2

3.4 International and National status

Most of the International researchers have established the influence of several factors on the quality of recording of speech samples. They have also compared the recordings obtained through different devices. But none of them have arrived at an optimal configuration combining microphone, digitizing system as well as processing system. There is no reported evidence on the National status.

2

3.5 Importance of the proposed project in the context of current status

Closeness of speech samples recorded for analysis to the original ones, play a crucial role in research as well as diagnosis and rehabilitation of persons with communication disorders. Considering the significant difference in the quality and reliability between different methods for voice and speech recording, it is essential to arrive at a recording set up which provides least variation in comparison with the bench mark. The proposed project is to establish a cost effective and easily available recording configuration and thus would ensure proper recording of speech samples.

2

4.0 Work plan

4.1 Method

Subjects / Participants

Ten normal healthy individuals (Five males and five females) between 21 and 36 years of age who are native Kannada speakers will participate in the study. It will be ascertained from a structured interview that the participants doesn't have any neurological, psychological or voice disorders and are nonsmokers and nonalcoholic. Only those without any hearing related problem like tinnitus etc. will be recruited in to the study. Another group of six normal healthy individuals (three males and three females) satisfying the same inclusion and exclusion criteria will participate in the validation process.

Material

7

Voiced (/b/, /d/, /g/) and unvoiced (/p/, /t/, /k/) stop consonants paired with vowel /a/ will be used as the speech stimuli. Speech stimuli uttered by each participant will be recorded during one session for all the combinations, with rest time in between. Participants were provided with recording microphone, seated in a sound treated room, and instructed to read stimuli from a printed paper. The speakers read all the stimuli aloud in a "normal, conversational voice". Participants will be instructed to utter each stimuli three times, with pauses after each utterance. An inter stimulus interval of six seconds will be maintained

between each syllable. Sixty different combinations of recording set up will be set up with ten laptops of different configuration, each one with a different sound card, with different processing features and six microphones (that can be coupled with laptops) with different characteristics as shown in Table – 1 below:-

Table – 1: Configuration of Laptop, sound card and microphone.

Laptop configuration		Sound card specifications		Microphone type	
L1	Intel Core i7-4110 CPU 2.00 GHZ, with Windows 8.1, 64 bit OS and 4GB RAM	S1	24bit, 192000Hz	M1	In built microphone of the laptop
L2	Intel Core i5-6200 CPU 2.3 GHZ, with Windows 8.1, 64 bit OS and 4GB RAM	S2	24bit, 96000Hz	M2	External Mic Creative HS 150
L3	Intel Core i5-6200 CPU 2.3 GHZ, with Windows 8, 32 bit OS and 4GB RAM	S3	24bit, 88200Hz	M3	External Mic Creative HS 390
L4	Intel Core i5-5200 CPU 2.2 GHZ, with Windows 8.1, 64 bit OS and 8GB RAM	S4	24bit, 48000Hz	M4	External Mic i ball rocky
L5	Intel Core i5-2430 CPU 2.4 GHZ, with Windows 8, 64 bit OS and 4GB RAM	S5	24bit, 44100Hz	M5	External Mic Frontech Jil-3442
L6	Intel Core i5-4200 CPU 1.6 GHZ, with Windows 7, 32 bit OS and 4GB RAM	S6	16bit, 192000Hz	M6	USB External Mic Creative HS 720
L7	Intel Core i5-2450M CPU 2.50 GHZ, with Windows 8 PRO, 64 bit OS and 8GB RAM	S7	16bit, 96000Hz		
L8	Intel Core i3-2350 CPU 2.30 GHZ, with Windows 7, 64 bit OS and 2GB RAM	S8	16bit, 88200Hz		
L9	Intel Core i3-M 380 CPU 2.53 GHZ, with Windows 7, 32 bit OS and 4GB RAM	S9	16bit, 48000Hz		
L10	Intel Core i3-2348M CPU 2.3 GHZ, with Windows 8.1, 32 bit OS and 2GB RAM	S10	16 bit, 44100Hz		

Procedure

Speech stimuli uttered by the participants will be recorded with sixty different combinations of recording set up as shown in Table – 2 below:-

Table – 2: Recording configurations

Recording combination	Number of situations
Laptop configuration L1 with 1 sound card S1 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L2 with 1 sound card S2 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L3 with 1 sound card S3 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L4 with 1 sound card S4 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L5 with 1 sound card S5 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L6 with 1 sound card S6 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L7 with 1 sound card S7 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L8 with 1 sound card S8 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L9 with 1 sound card S9 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Laptop configuration L10 with 1 sound card S10 paired with microphones M1, M2, M3, M4, M5 and M6 respectively.	6 Nos.
Bench mark system	1 No.
Total Number of recording situations	61 Nos.
Total Number of recorded speech samples	61 Nos. x 6 stimuli = 366 Nos. 366 Nos. x 10 subjects = 3660

The microphone of the recording set up will be kept at a distance of 15 cms from the mouth of the participant for all combinations. The same set of stimuli will also be recorded by the bench mark system.

The estimated parameters will be compared with those of the recoding obtained using a bench mark configuration. The bench mark configuration involves a high quality recording system coupled with a precision condenser microphone.

Analyses

The following analyses will be carried out using Adobe Audition and Praat:-

- a) Measurement and comparison of VOT, transition and closure duration in recorded speech of all combinations including benchmark system and original speech.
- b) Measurement and comparison of Fundamental frequency and formants - First, Second and Third in recorded speech of all combinations including benchmark system and original speech.
- c) Estimation and comparison of spectral tilt in recorded speech of all combinations including benchmark system and original speech.

Appropriate statistical analysis will be done to identify the optimal configuration. The identified optimal configuration will be validated with the recorded speech samples of another group of participants.

3

6.0 Implications of the results of the study (Illustrative)

- a) Presentation of scientific papers in professional seminars / publication of articles

Research findings will be presented in National / International seminars and also would be published in National / International journals.

- b) Discussion with professionals

Presentation of papers resulting from the study in National / International seminars / conferences will enable discussion with other professionals.

c) To utilize the results in the development of remediation

Results of the study will provide guidelines to the researcher and clinician while recording the speech samples.

2
7.0

Utilization of results of the study

Outcome of the study will be a cost effective, widely available and user friendly configuration for recording speech samples with optimum quality. Optimum quality recording will in turn lead to accurate assessment and diagnosis which will result in effective rehabilitation. Accurate recording will also lead to reliable research findings while using these recorded samples for analysis.

Development of a cost effective configuration for optimal quality recording for speech analysis

ORIGINALITY REPORT

13%

SIMILARITY INDEX

8%

INTERNET SOURCES

5%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Li, Bo Qiu, Bo Lee, David S.M. Walton, Z. "Fructose-1,6-bisphosphatase opposes renal carcinoma progression.(RESEARCH: LETTER)", Nature, Sept 11 2014 Issue
Publication 4%
- 2 www.aiishmysore.in
Internet Source 4%
- 3 Submitted to All India Institute of Speech & Hearing
Student Paper 2%
- 4 www.sony.com.my
Internet Source 1%
- 5 www.computronestore.com
Internet Source 1%
- 6 Vogel, Adam P., and Angela T. Morgan. "Factors affecting the quality of sound recording for speech and voice analysis", International Journal of Speech-Language Pathology, 2009. 1%

7

Li, Shu-Chen, Susanne Passow, Wilfried Nietfeld, Julia Schröder, Lars Bertram, Hauke R. Heekeren, and Ulman Lindenberger.

"Dopamine modulates attentional control of auditory perception: DARPP-32 (PPP1R1B) genotype effects on behavior and cortical evoked potentials", *Neuropsychologia*, 2013.

Publication

1%

8

letsforum.com

Internet Source

<1%

9

www.paphoscomputers.com

Internet Source

<1%

EXCLUDE QUOTES ON
EXCLUDE ON
BIBLIOGRAPHY

EXCLUDE MATCHES < 7 WORDS