

**Effects of Microphone Type and Distance  
Upon the Spectra of Speech Sounds**

by

Dameon C. Harrell

Submitted to the Department of  
Electrical Engineering and Computer Science in  
Partial Fulfillment of the Requirements for the  
Degrees of

Bachelor of Science and Master of Engineering in  
Electrical Engineering

at the

Massachusetts Institute of Technology

May 21, 1999

June 1999

© 1999 Dameon C. Harrell

All rights reserved.

The author hereby grants to MIT permission to reproduce and to  
distribute publicly paper and electronic copies of this thesis document in whole or in part  
and to grant others the right to do so.

Signature of Author .....

Department of Electrical Engineering and Computer Science

May 21, 1999

Certified by .....

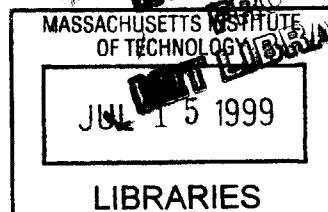
Kenneth N. Stevens

Clarence J. Lebel Professor Of Electrical Engineering

Thesis Supervisor

Accepted by .....

Arthur C. Smith  
Chairman, Department of Electrical Engineering and Computer Science on Graduate Theses



ENG

Effects of Microphone Type and Distance  
Upon The Spectra of Speech Sounds

by  
Dameon Harrell

Submitted to the Department of Electrical Engineering and Computer Science  
In Partial Fulfillment of the Requirements for the Degrees of  
Bachelor of Science in Electrical Engineering and Computer Science and  
Master of Engineering and Computer Science

## **ABSTRACT**

This thesis examines and attempts to quantify and explain the effects of microphone type and placement upon the spectra of selected speech sounds using theories of speech production, acoustics, and microphone characteristics. Three vowels, the voiced stop consonants, as well as two nasal consonants were chosen to be studied for the thesis research. The research was conducted by having several speakers say utterances containing the selected speech sounds while being recorded by two different types of microphones simultaneously. Trends in the differences between the spectra of each microphone were found for each speaker's vowels, nasal consonants, and the voiced stop consonants. However, each of these trends were different from each other in ways that were both expected and unexpected. The results imply that more detailed study of the acoustic nearfield about the head is necessary in order to obtain a better understanding of the differences that occurred.

Thesis Supervisor: Kenneth N. Stevens  
Title: Clarence J. Lebel Professor of Electrical Engineering

# Table of Contents

<b>1</b>	<b>Introduction and Theory</b>	<b>6</b>
1.1	Introduction	6
1.2	Theory	8
<b>2</b>	<b>Setup and Procedures</b>	<b>15</b>
2.1	Speakers	15
2.2	Equipment and Connections	15
2.3	Setup and Procedure	16
2.4	Data Extraction	17
2.4.1	Vowel Spectra	18
2.4.2	Voiced Stop Consonant and Nasal Consonant Spectra	18
2.4.3	Measuring the differences in the spectra	19
<b>3</b>	<b>The Vowels</b>	<b>20</b>
3.1	Results	20
3.1.1	[a]	20
3.1.2	[i]	24
3.1.3	[u]	28
3.2	Analysis	31
<b>4</b>	<b>Voiced Stop Consonants</b>	<b>34</b>
4.1	Results	34
4.1.1	[b]	35
4.1.2	[d]	38
4.1.3	[g]	41
4.1.4	Comparison of each Voiced Stop Consonant across the vowels	44
4.1.5	Comparison of the Voiced Stop Consonants in the context of [a]	47
4.2	Analysis	48
<b>5</b>	<b>Nasal Consonants</b>	<b>49</b>
5.1	Results	49
5.1.1	[m] and [n] from bama and bana	50
5.1.2	[m] and [n] from bimi and bini	54
5.1.3	[m] and [n] from bumu and bunu	57
5.1.4	Individual speaker results across all [m] and [n]	60
5.1.5	Interspeaker [m] and [n] comparison	62
5.2	Analysis	65
<b>6</b>	<b>Conclusions and Recommendations</b>	<b>65</b>
	<b>Acknowledgments</b>	<b>70</b>
	<b>References</b>	<b>71</b>

## List of Figures

1.1	Close-talking microphone position with respect to the mouth	7
1.2	Directivity pattern of close-talking microphone	9
1.3	Close-talking microphone's frequency response	10
1.4	Measures used for dimensions of the lips	12
1.5	Angle of approach for sounds radiating from the neck	13
2.1	list of utterances	15
3.1	am [a] difference plot	20
3.2	ks [a] difference plot	21
3.3	ss [a] difference plot	22
3.4	[a] difference plot; comparison of all speakers	23
3.5	am [i] difference plot	24
3.6	ks [i] difference plot	25
3.7	ss [i] difference plot	26
3.8	[i] difference plot; comparison of all speakers	27
3.9	am [u] difference plot	28
3.10	ks [u] difference plot	29
3.11	ss [u] difference plot	30
3.12	[u] difference plot; comparison of all speakers	31
4.1	[b] from aba difference plots for all speakers	35
4.2	[b] from ibi difference plots for all speakers	36
4.3	[b] from ubu difference plots for all speakers	37
4.4	[d] from ada difference plots for all speakers	38
4.5	[d] from idi difference plots for all speakers	39
4.6	[d] from udu difference plots for all speakers	40
4.7	[g] from aga difference plots for all speakers	41
4.8	[g] from igi difference plots for all speakers	42
4.9	[g] from ugu difference plots for all speakers	43
5.1	am [m] and [n] difference plots for bama and bana	50
5.2	ks [m] and [n] difference plots for bama and bana	51
5.3	ss [m] and [n] difference plots for bama and bana	52
5.4	am [m] and [n] difference plots for bimi and bini	54
5.5	ks [m] and [n] difference plots for bimi and bini	55
5.6	ss [m] and [n] difference plots for bimi and bini	56
5.7	am [m] and [n] difference plots for bumu and bunu	57
5.8	ks [m] and [n] difference plots for bumu and bunu	58
5.9	ss [m] and [n] difference plots for bumu and bunu	59
6.1	Sketch of general trends for each phone type	65
6.2	Near field response about a coronal cross-section of the head for three vowels	68

## List of Tables

4.1	[b] 1st and 2nd harmonic differences across vowels	44
4.2	[d] 1st and 2nd harmonic differences across vowels	45
4.3	[g] 1st and 2nd harmonic differences across vowels	46
4.4	Voiced Stop Consonant measurements in the vowel context of [a]	47

# 1 Introduction and Theory

## 1.1 Introduction

The goal of the research discussed in this thesis was to examine and attempt to quantify and explain the observed difference in spectral characteristics of particular speech sounds as recorded by two different types of microphones. More specifically, these particular speech sounds were recorded simultaneously on two separate channels by an omnidirectional microphone and a head-set mounted super-cardioid close-talking microphone, and spectra from the resulting signals were generated. The differences between the amplitudes of the spectral peaks (harmonics) were measured and attempts were made to account for these differences with models of speech production and acoustics in general, and microphone specifications such as response and directivity pattern. This thesis attempts to determine if the type of spectral differences that are predicted are generally what are observed when the two recordings are examined.

The research conducted for this thesis is relevant to applications such as speech and speaker recognition which are dependent upon the spectral characteristics of speech sounds as they are recorded by the microphones being used for their respective applications. The results yielded by the research could aid in making advancements in the creation of more robust recognition software such that the performance of the system is not microphone dependent.

Currently a variety of dictation software is available, and this software performs with high accuracy if it is used correctly. A speech recognition system of this type is very constraining with respect to the environment it must be used in and equipment it must be used with. For example, if

a user attempts to use a type of microphone different from the type for which it was trained, the system will work less effectively.

The differences in the spectra of speech sounds transduced by an omni-directional and a directional, close-talking microphone can be significant. This is especially true for those phones for which the primary acoustic cues as transduced by the close-talking microphone are sensitive to the distance from and/or the direction of approach of the propagating wavefront with respect to the microphone. The possibility that the reception of an acoustic cue of a phone will be affected by the differences in distance and placement of the microphones, with respect to the source of the sound, increases for those consonantal sounds that have multiple sources of sound radiation or sources other than the mouth opening. Figure 1.1, given below, shows a sketch of how the headset with a close-talking microphone is worn. It also displays the microphone's position with respect to the mouth.



**Figure 1.1:** Close-talking microphone position with respect to the mouth

In order for the results from the spectral comparison experiment to be meaningful, it was necessary to carefully control the environment. More precisely, ambient sounds and noise from the equipment (preamplifiers and connections) had to be minimal with respect to the desired speech signals. In addition, it was desirable for the channels for each microphone to have equivalent or near equivalent responses so that differences in the recorded signals could be accurately attributed to the differences in microphones. It was expected that a pattern would be observed in the differences of the spectra of specific phones for each speaker and between speakers, and that these differences could be readily accounted for by theory, as summarized in the following section.

## **1.2 Theory**

When sound is radiated exclusively from an opening of the mouth, as it is during the production of a vowel, the signal recorded by the two different types of microphones is expected to be very similar. Nevertheless, distinct differences between the signals could be present depending on the size and shape of the mouth opening. Because close-talking microphones are positioned in the near field of the source at the mouth, divergence from “normal” acoustic behavior can be expected for different mouth opening shapes. The sound originating from the mouth will not be attenuated at 6 dB per doubling of distance as it approaches the microphone as it would if the microphone was positioned in the far field.

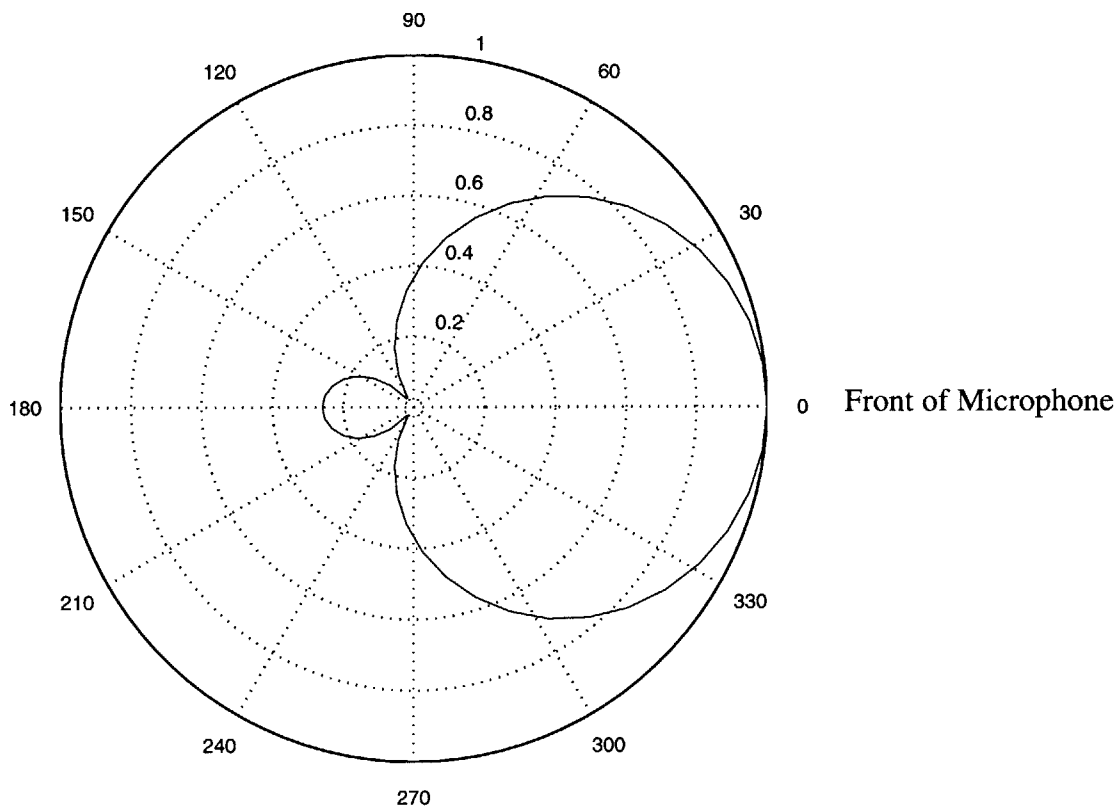
When the sound is radiated from the surface of the neck (in the region of the larynx) or the nose the signals recorded by each microphone will be different. These expectations are due in part to the fact that the directivity pattern of the close talking microphone which was used in this experi-



ment is a supercardioid and in part to the differences in distance from the close-talking microphone to the mouth, nose, and neck. The equation for the directivity response is:

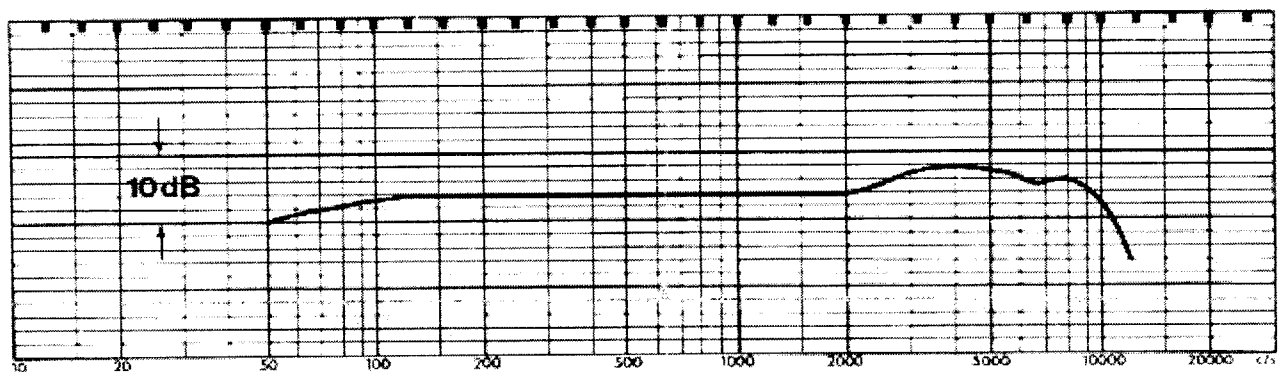
$$\rho = 0.37 + 0.63 \cos\theta [1]$$

A diagram of the response as a function of angle (directivity pattern) is given in Figure 1.2. The response at plus or minus 90 degrees (relative to the front of the microphone, at zero degrees) is about -8.6 dB, or the sound is attenuated by a factor of 0.37 relative to the gain at the front of the microphone which is 1. The response at 180 degrees is -11.7dB. [1]



**Figure 1.2:** Directivity pattern of close-talking microphone

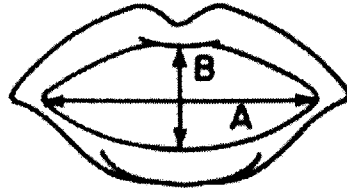
The outer most circle in the diagram of Figure 1.2 represents the contour of total pickup, or 1. Each inner circle decreases the fraction of pickup by 0.2 and the very center of the diagram represents 0, or no pickup. The no pickup angle is about 126 degrees. The microphone's response as a function of frequency is given in Figure 1.3



**Figure 1.3:** Close-talking microphone's frequency response [2]

Certain vowels with different size mouth openings were selected to demonstrate situations in which the signal was predicted to look similar. The vowel [a] was chosen because of its lowered mandible and relatively large mouth opening, while [u] was chosen because it is pronounced with rounded, protruding lips and a very small mouth opening. Finally, the vowel [i] was selected because it is realized with the lips spread apart and a mouth opening of a size that is between the two aforementioned vowels. In addition, the different vowels were used to present each of the featured consonants within different phonemic contexts.

Initially, it was thought that [u] would demonstrate the smallest differences in the spectra derived from the signals of the two different microphones. The reason for this thinking is that the small mouth opening for [u] would cause the wave propagating from it to appear as if it were radiating from a simple source, at least over the range of frequencies being examined for this thesis (0 to 4kHz). The average size mouth opening during the articulation of [u] is less than 1cm in diameter [3]. The wavelength at 4kHz is about 8.6 centimeters so the mouth opening dimension is less than eighth of a wavelength, so the radiating sound should appear to be emanating from a simple source and should radiate equally in all directions (except for the baffling effect of the head). The mouth opening for [i] has its largest dimension along segment B in Figure 1.4 and it is on the order of 4 cm to 4.5 cm [3]. The sound wave which propagates from the mouth during the articulation of [i] will not look like it is coming from simple source for frequencies over, approximately, 2kHz. The mouth opening to produce the vowel [a] has long A and B dimensions (Figure 1.4). However, neither of these dimensions is any larger than the largest dimension for [i], so its propagation is expected to exhibit simple source behavior over the same frequency range as [i]. But, because it has such a large mouth opening and due to the proximity and location of the close-talking microphone it is figured to display the most different spectral characteristics between the two microphones of the all vowels. Since the close-talking microphone is off to the side and not in front of the mouth, the part of the sound wave, in the higher frequencies, travelling from the side of the mouth furthest from the microphone will not necessarily be in phase with the sound coming from the near side and thus possibly causing interference. This is also true for [i].

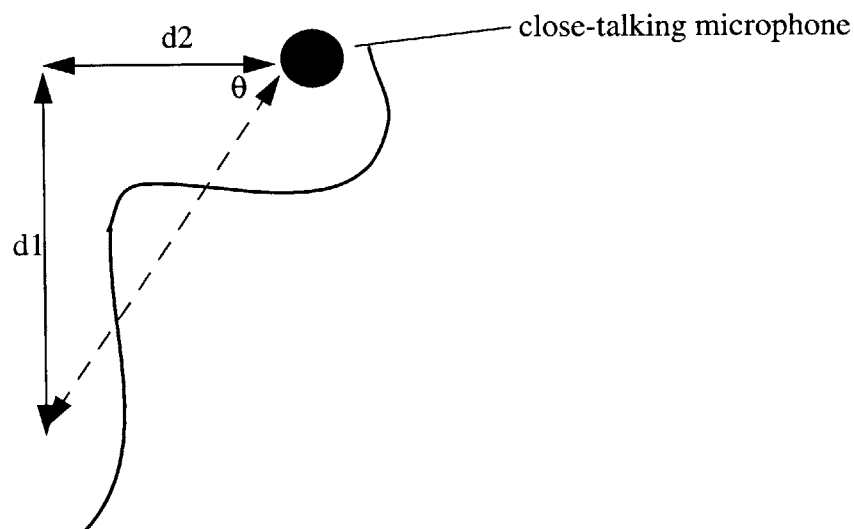


**Figure 1.4:** Measures used for dimensions of the lips [3]

The consonants chosen for study were the voiced stops and nasals [m] and [n]. These were selected because they were thought to be the phones that would show the most significant difference between the signals recorded by the two different types of microphones.

The voiced stop consonants are /b/, /d/, and /g/. The type of microphone and its position in relation to a speaker's mouth is important to the spectral characteristics of these consonants because of the voicing before their release burst. It is expected that this pre-voicing radiates mostly from the front surface of the neck because the mouth and nasal cavities are closed. Therefore, the signal received by the two types of microphone should be very different. Figure 1.5 shows that the angle at which the wavefront from the neck surface approaches the close-talking microphone can never be greater than 90 degrees, but will probably be greater than 45 degrees because the sound is coming from a point of the neck surface which is further down than neck ( $d_1$ ) than the horizontal distance from the neck surface area of greatest vibration to the microphone ( $d_2$ ). The angle  $\theta$  increases as dimension  $d_2$  increases. So, a valid assumption would be that the maximum attenuation due to the microphone directivity is -8.6 dB. Therefore, theoretically, it is believed that the spectra of signals from the closetalking microphone would be approximately 5 to 9 dB down from

those transduced by the omnidirectional microphone if the signals were normalized to account for the attenuation due to the difference in distance travelled by the signal from the close-talking microphone to the omnidirectional microphone.



**Figure 1.5:** Angle of approach for sounds radiating from the neck

The research also focused on the English nasal consonants [m] and [n]. These are produced with a closed mouth and open nasal cavity. As a result there is little to no pressure build up and the majority of the sound emanates from the openings of the nose. The nasals were chosen for this reason and also because they may have a contribution of sound radiating from the neck surface. The difference between [m] and [n] is mainly perceived upon release. Therefore, it is anticipated that both [m] and [n] will exhibit very similar differences in the spectra produced by the two different types of microphones. Because of the supercardioid directivity pattern of the close-talking microphone sound propagating from both sources could be attenuated with respect to sound arriving normally to the front of the microphone. Consequently, very different spectra should be produced by the analysis of the signal recorded by each microphone type.

It is difficult to estimate what this difference might be because a significant amount of sound may be radiating from the neck surface as well as the nose, and they may interact in such a way that the signals are not additive because their sources are so close and coupled. And while this interaction may be less of a factor for the close-talking microphone, the signals from nose and neck will definitely interact before reaching the omnidirectional microphone possibly leading to constructive or destructive interference of the acoustic signal. This interaction would affect the higher frequency components of the sound much more than the lower frequency ones. Also, there could be physical interference from the chin and jaw obstructing the path of the sound wave propagating from the neck. So it is expected that there may be some divergence in the similarity of the higher frequency region of the nasal spectra produced by the two different microphones. In general, the nasal consonant spectra from the close-talking microphone should have less energy than the spectra from the omni-directional microphone, if the microphone outputs are normalized for the vowels.

## 2 Setup and Procedures

### 2.1 Speakers

Three speakers, two males (am and ks) and one female (ss), all students or faculty/staff from MIT's Speech communication group, were asked and agreed to be recorded speaking a set of nonsense utterances which contained all of the previously mentioned vowels and consonants which were to be studied. The list of nonsense utterances is given below in Figure 2.1

bama  
bana  
bimi  
bini  
bumu  
bunu  
aba  
ada  
aga  
ibi  
idi  
igi  
ubu  
udu  
ugu

**Figure 2.1:** list of utterances

### 2.2 Equipment and Connections

The utterances were recorded using an Electrovoice dynamic, omnidirectional microphone and a Sennheiser pressure-gradient, HMD410 Headset microphone (the close-talking microphone). Each microphone was connected to a Shure microphone mixer (both mixers have a flat frequency response + or - 3dB from 40Hz to 20kHz [4]) which served as a preamplifier. The omnidirec-

tional microphone was connected to the Shure M67 Microphone mixer through the XLR microphone input. The headset microphone was connected to a Shure M68 microphone mixer via an XLR microphone input with the aid of a 1/4" phone jack to XLR 3-pin male connector transformer. The headphone phone jack output of the M67 was connected to a 1/4" RCA cable, which was in turn connected to a female to female RCA coupler that was connected to one line of an RCA to stereo mini cord. The other line of the RCA to stereo mini cord was connected directly to the RCA line output of the M68 mixer. The mini plug of the RCA to mini chord was connected to the line input jack of a Sony TCD-D8 stereo DAT recorder. The RCA to mini cord was connected such that the close-talking microphone signal was in the left channel of the DAT recorder, and that of the omnidirectional in the right channel.

The signal recorded on the DAT was down sampled and the two channels put in separate files using the Digidesign SoundDesigner audio editing software on a Macintosh II ci. These files were then analyzed with the aid of XKL, a speech and spectral analysis software program.

The recording was done in a quiet, acoustically dampened room.

## **2.3 Setup and Procedure**

Each speaker was recorded simultaneously by both microphones while saying the list of utterances. The speaker wore the headset and adjusted the head strap and microphone boom until he or she was comfortable, the microphone usually ending up near the right corner of the lips. Next, he or she was asked to stand about six inches in front of the omnidirectional microphone which was suspended from the ceiling with the cord slug over a taut string which was drawn across the



ceiling. Six inches was the chosen distance because of the desire to eliminate any reflections or the room acoustics issues from affecting the signals. The trade-off was that now the omnidirectional microphone was in the nearfield of the sounds. The height of the omnidirectional microphone was adjusted so that it hung at about the level of each speaker's mouth. Once the speaker was comfortable, the levels of the microphones were adjusted to try to obtain equal levels on both channels while avoiding clipping by using the DAT's level meter. Then they were recorded speaking the list of utterances. The DAT recorder was set to sample at a 44.1kHz sampling rate. After being recorded and sampled the digital speech signals were uploaded into the Macintosh where the SoundDesigner software was used to down sample the signal twice to a sampling rate of 11,025 Hz. Next the downsampled signal from each channel was put in a separate aiff file (apple's audio file format) and transferred as raw data to the Speech Communication Group's Linux computers. Once the files were in the Linux system, they were converted to wav files, at which point they could be parsed into the separate utterances and each utterance analyzed separately for each microphone with the XKL software.

## **2.4 Data Extraction**

The XKL software was used to create spectra at various points in each waveform to extract information. The window size was adjusted to fit the pitch of speaker's voice such that it contained between three and four pitch periods. The window was centered over one of the pitch periods to achieve maximum harmonic resolution. Two recordings of each utterance were created, one by the omnidirectional microphone and the other by the close-talking microphone. Spectra from the same point in time of each waveform as captured by the two recordings of an utterance were needed for analysis. These were obtained by using landmarks in the waveforms as well as taking

into account the delay time between the signal of the two microphones. The windows used for the waveforms of both recordings of an utterance were identical.

### **2.4.1 Vowel Spectra**

To obtain spectra of each vowel from each speaker, the utterances aba, ibi, and ubu were used. For each vowel two spectra were made by windowing at two places in the middle of the vowel in the first vowel, before the consonant, and two places in the middle of the vowel following the consonant. To measure the differences in the spectra that were due to microphone type and placement, the spectra were normalized such that the first formant peaks were equalized, or, in other words, their difference was made zero.

### **2.4.2 Voiced Stop Consonant and Nasal Consonant Spectra**

The spectra used for finding the differences in the spectra of each of the stop consonants from the two different microphones were obtained in a similar manner. The signal was normalized to account for attenuation due to distance traveled by equalizing the amplitude of the first formant peaks of the spectra derived from the middle of the preceding vowel.

The nasals were handled in a similar manner, except the middle of the preceding vowel was not used. Instead, the part of the vowel immediately following the initial [b] in each of the nasal utterances was used to avoid windowing a portion of the waveform where the vowel may have been nasalized.

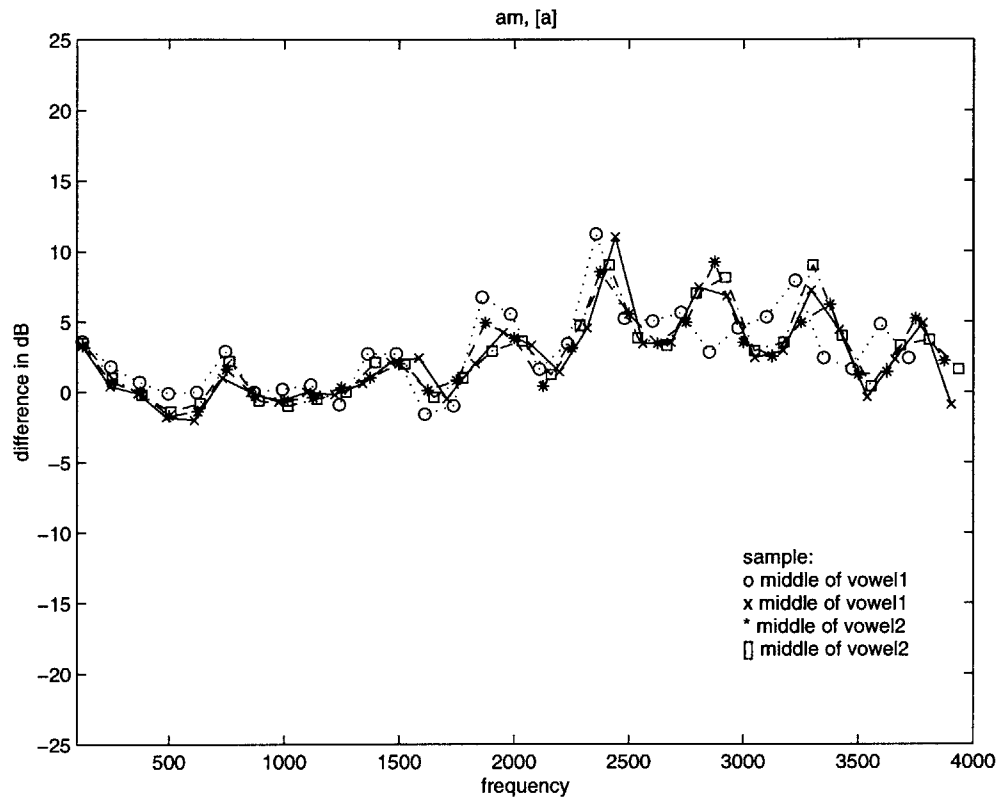
### 2.4.3 Measuring the differences in the spectra

The XKL software has a feature which will automatically select the harmonic peaks of the spectra in the spectra window, and record the amplitudes of these peaks. This software was used to record the harmonics' amplitude values from spectra taken from the same point in time of each of the two recordings of an utterance. These values were put into a vector in a matlab file. Each of the two recordings had separate vectors for the same set of harmonic amplitude values from their "matching" spectra. The vector of the omni-directional microphone was subtracted from the vector of the close-talking microphone to form a difference vector. If a harmonic of the spectra from the omnidirectional microphone was of lesser amplitude than that of the same harmonic in the spectra of the close-talking microphone, that harmonic in the difference vector would have a positive value. This difference vector was plotted against a frequency vector which contained the frequencies corresponding to the harmonics of the differences contained in the elements of the difference vector. The frequency vector was created by multiplying a harmonic vector, a vector containing the number of each harmonic in the difference vector, by the fundamental frequency of the corresponding spectrum. For example, if only the values of the first 5 harmonics were recorded and the fundamental frequency of the corresponding spectra was 100 Hz, the frequency vector,  $f$ , would be defined as  $f = 100 \times [1 \ 2 \ 3 \ 4 \ 5] = [100 \ 200 \ 300 \ 400 \ 500]$ . Some of the difference plots do not contain a continuous set of data points because they are missing the value of the differences at some frequencies, due to noise in the spectra which made the values of the harmonics at those frequencies unreliable. Both of the spectra needed clean data in order to obtain reliable differences. If one of the spectra was noisy in a particular region no data could be used from that region even if the other spectrum was uncontaminated by noise.

# 3 The Vowels

## 3.1 Results

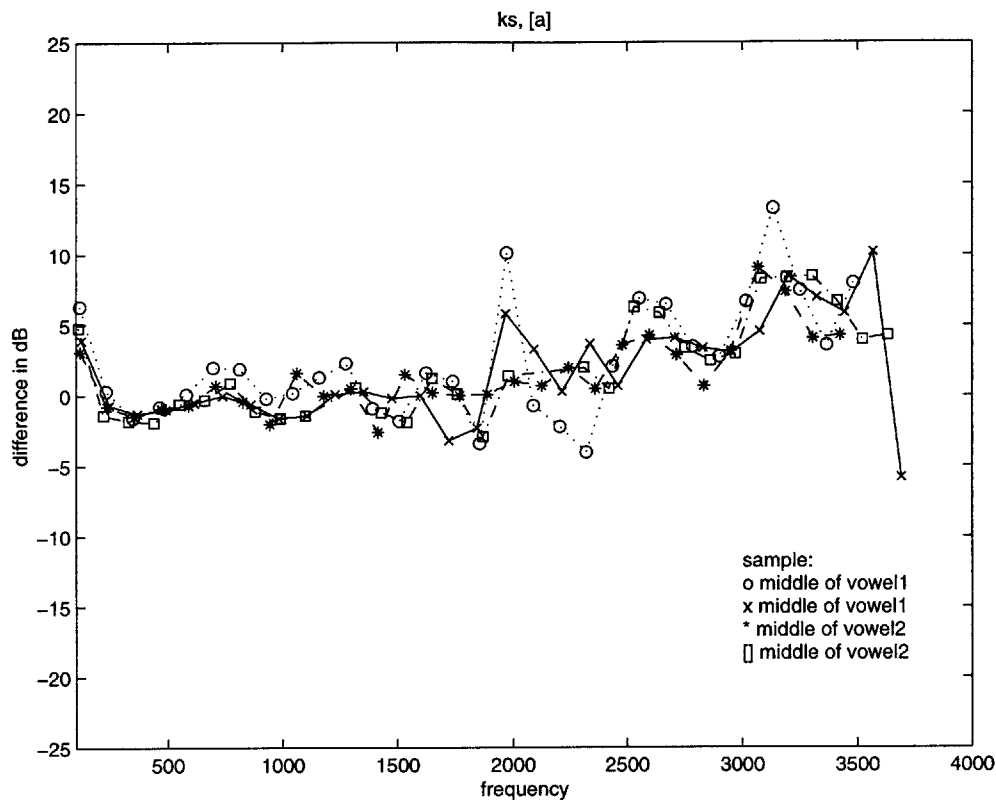
### 3.1.1 [a]



**Figure 3.1:** am [a] difference plot

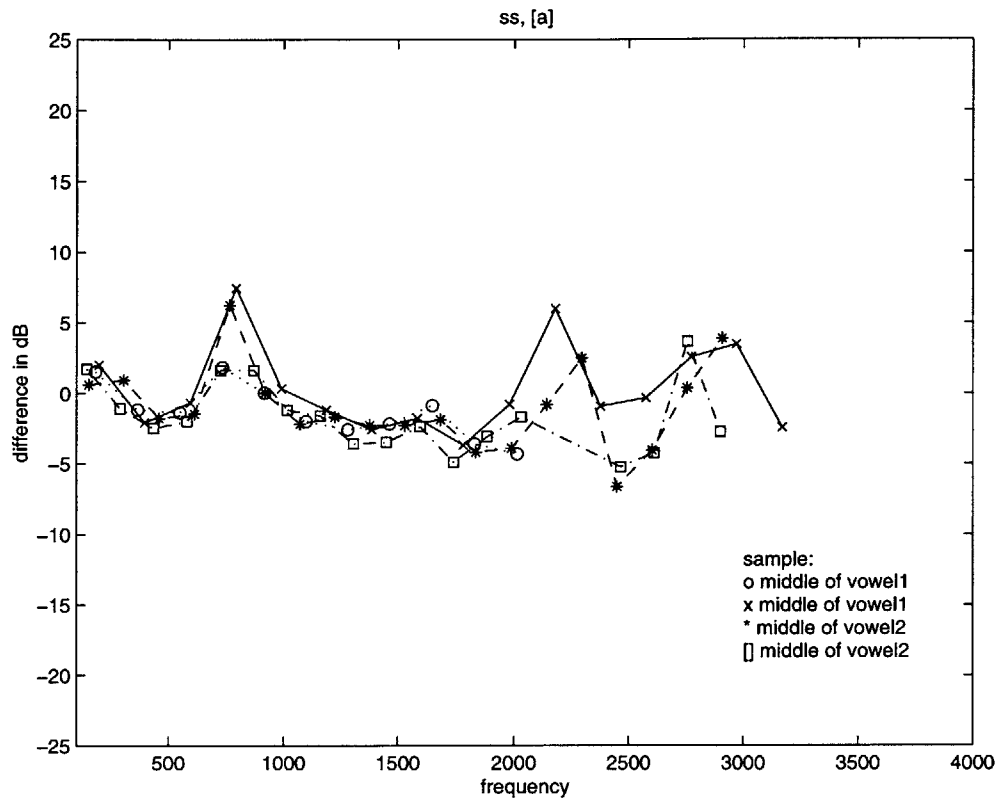
Figure 3.1 contains the difference plots from the four spectra of the vowel [a] from speaker am's utterance of aba. Each sample of the [a] difference plot tracks well with the others, but especially for 1500 Hz and below. Not only is the same pattern being followed in this region, but the values of the differences are nearly equivalent. In the higher frequencies the trajectory of each sample seems to diverge from the others by a small amount, nevertheless they all follow the same basic

pattern. However, beyond 1500 Hz the difference plot has peaks of greater than 5 dB in value whereas below 1500 Hz the difference remains within a plus or minus 4 dB range.



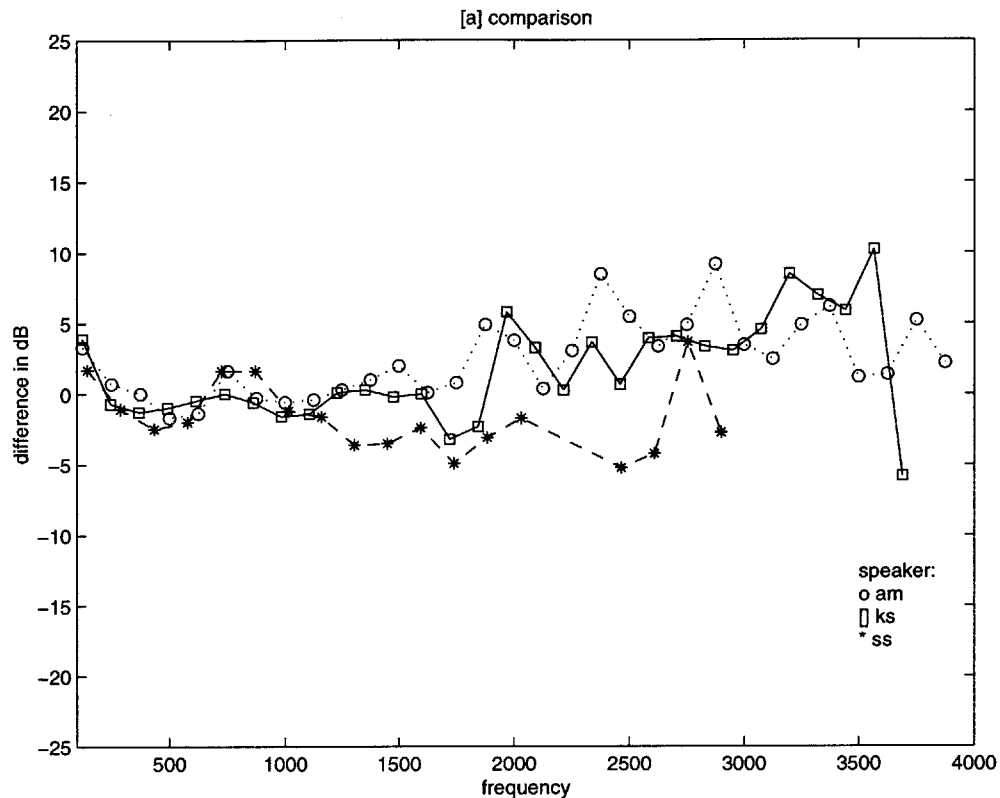
**Figure 3.2:** ks [a] difference plot

Figure 3.2 contains the same type of difference plots as those in 3.1 except these were derived from ks's version of the utterance aba. Again, there is a close tracking between each of the samples in the lower frequencies, although for ks they begin to diverge sooner than they did in am's plots, at about 1000 Hz. Beyond 1000 Hz the samples diverge in pattern, and also, as was the case for am, there are peaks of spectral differences up to 12 dB. Again, in the lower frequencies, 1000 Hz and below, the differences remain within plus or minus 5 dB and has a smooth pattern.



**Figure 3.3:** ss [a] difference plot

Figure 3.3 also contains difference plots for the vowel [a]; these data were extracted from the utterance aba spoken by speaker ss. Once again, as with am, the tokens for ss track closely up to about 1500 Hz. However, the exception from the previous two speakers is that the difference plots actually have a value greater than 5 dB for frequencies below 2000 Hz, with a peak at about 750 Hz. Other than this peak the pattern stays true to form with the differences staying within a 5 dB range and forming a smooth contour. Nevertheless, in the higher frequencies the tracking of the samples begin to diverge and have values of greater than 5 dB difference, although the differences for ss never exceed 10 dB as do those of the previous two speakers.

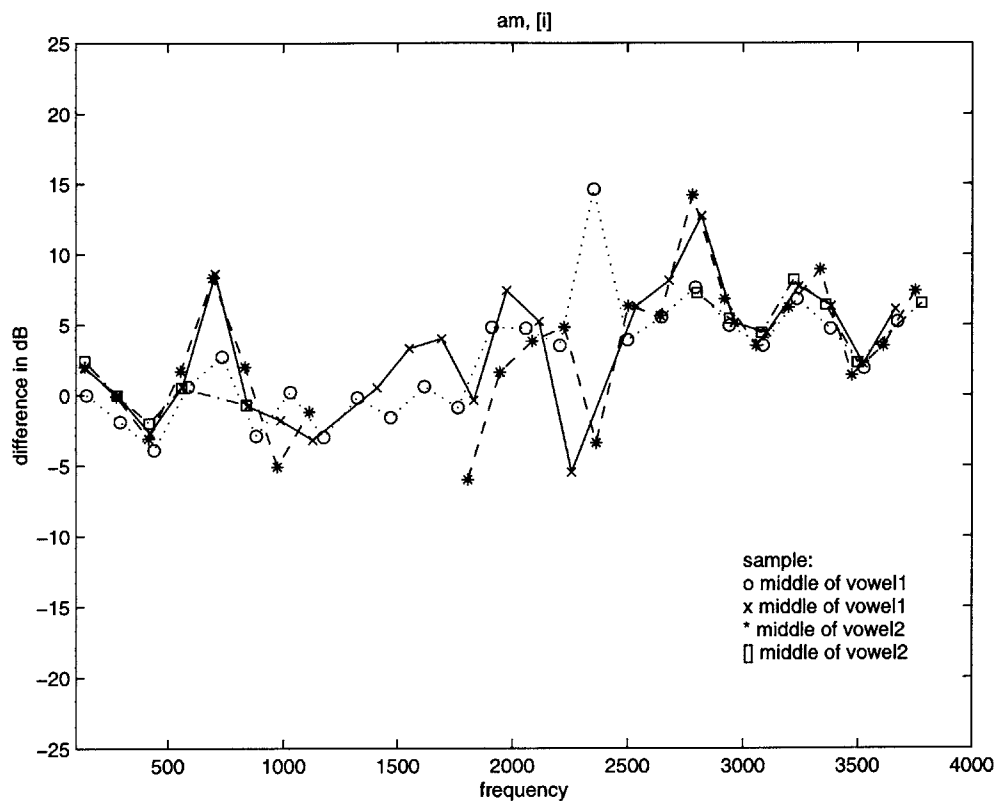


**Figure 3.4:** [a] difference plot; comparison of all speakers

Since each of the [a] samples of each speaker track so well in the lower frequencies and look nearly identical in the 100 Hz to 1000 Hz frequency range, one token from each can be used as a representative for that speaker within that frequency range. For the purposes of comparison the first sample of [a] for each speaker is shown in the plot in Figure 3.4. The differences of the [a] spectra as recorded by the different microphones track well for the different speakers through about 1000 Hz, and then the paths begin to diverge. All of the speakers share the common property of the first harmonic difference always being a positive value between 3 and 5 dB. This means that the close-talking microphone always has a greater first harmonic amplitude value than the omnidirectional microphone. After the first harmonic the differences tend to decrease and

become negative but never greater than a 5 dB absolute difference. Then the differences become positive and seem to shift between plus or minus 2 dB until the 1000 Hz point when they all begin to exhibit peaks and a general positive increase in difference. In the higher frequencies it can be expected then, that the close-talking microphone will exhibit higher harmonic amplitudes in its spectra.

### 3.1.2 [i]



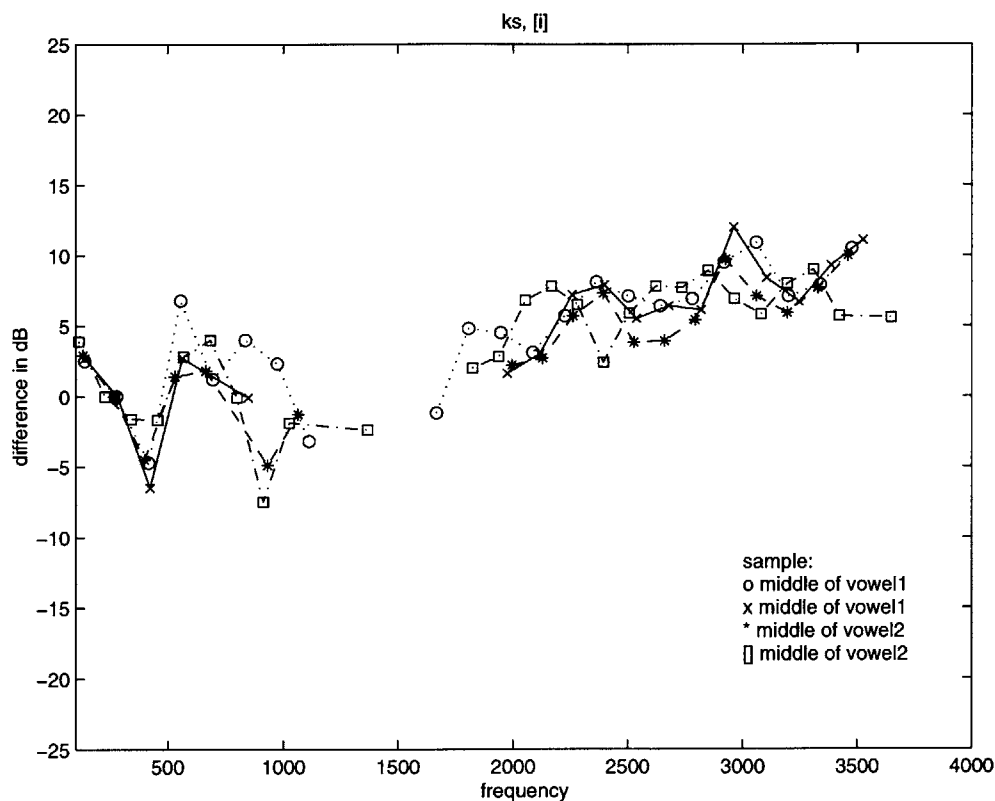
**Figure 3.5:** am [i] difference plot

Figure 3.5 displays the difference plots for am's samples of the vowel [i] in the utterance ibi.

Again, as with [a], there is a close tracking between the trajectories of the differences. However,



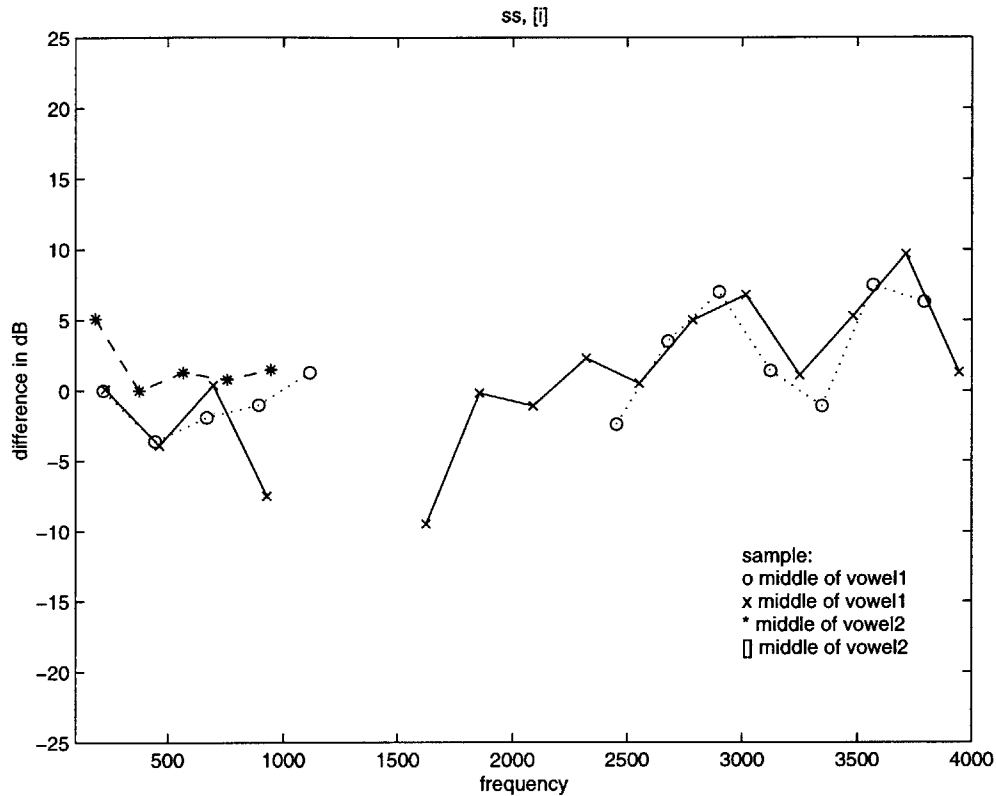
this time more discrepancies appear for frequencies above 1000 Hz. The peaks in difference in the 100 to 1000 Hz range occur at about 450 and 600 Hz, where the difference is as great as 10 dB. Between 1000 and 2000 Hz the differences seem to stabilize to within a plus or minus 5 dB range, but beyond 2000 Hz there are big difference peaks that are both positive and negative. However, at about 2750 Hz the samples seem to track well again and the differences range from no difference to about 7 dB.



**Figure 3.6:** ks [i] difference plot

The [i] difference plot samples for ks are shown in Figure 3.6 above. The same pattern that was present in the 100 to 1000 Hz frequency range for am can be seen in ks's plots as well. There is a large negative difference followed immediately by a large positive difference about 200 Hz later.

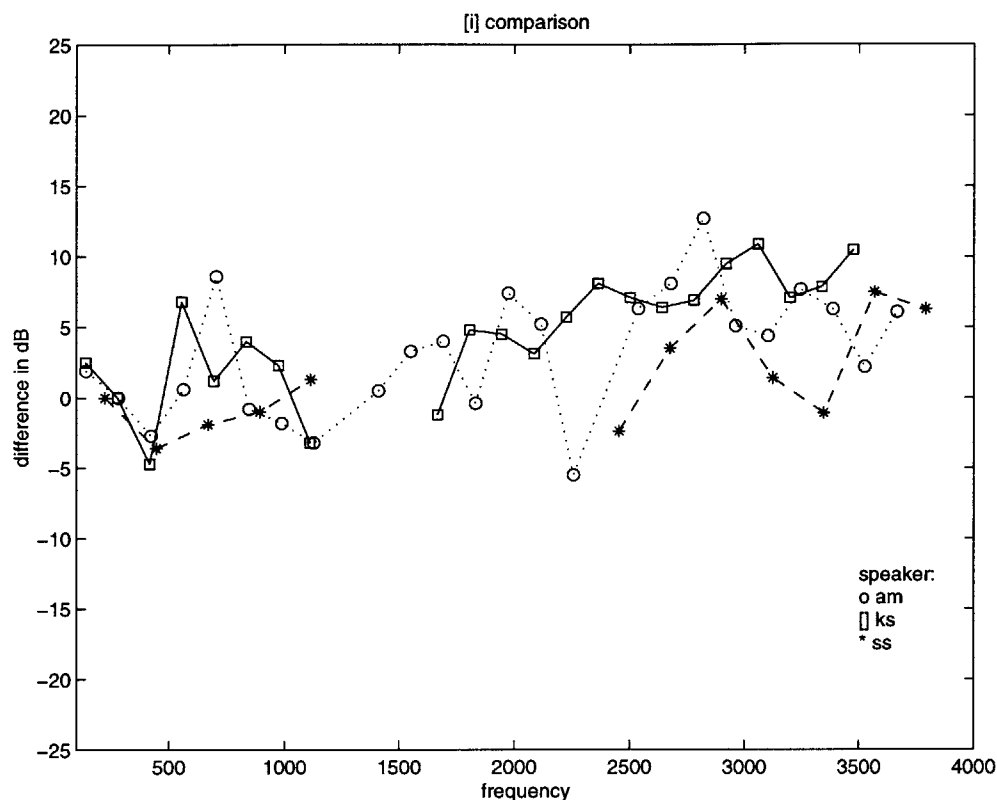
Because of noise there is not much data in the 1250 Hz to 2000 Hz frequency range. However, for frequencies greater than 2000 Hz the samples appear exhibit some degree of congruency but there is definitely a trend towards an increasing, positive difference.



**Figure 3.7:** ss [i] difference plot

It is difficult to make an assessment of what is occurring for the vowel of the speaker ss due to the lack of data. The second sample from the second vowel is not included because its spectrum was too noisy. Her vocal pitch is higher than the two male speakers so there are generally far fewer harmonics, but, due to noise, many differences were not reliable and therefore could not be included in the difference plot. From the information that is available in the difference plot in Figure 3.7 there appears to be the same dip in the differences around 450 Hz but there is no evidence

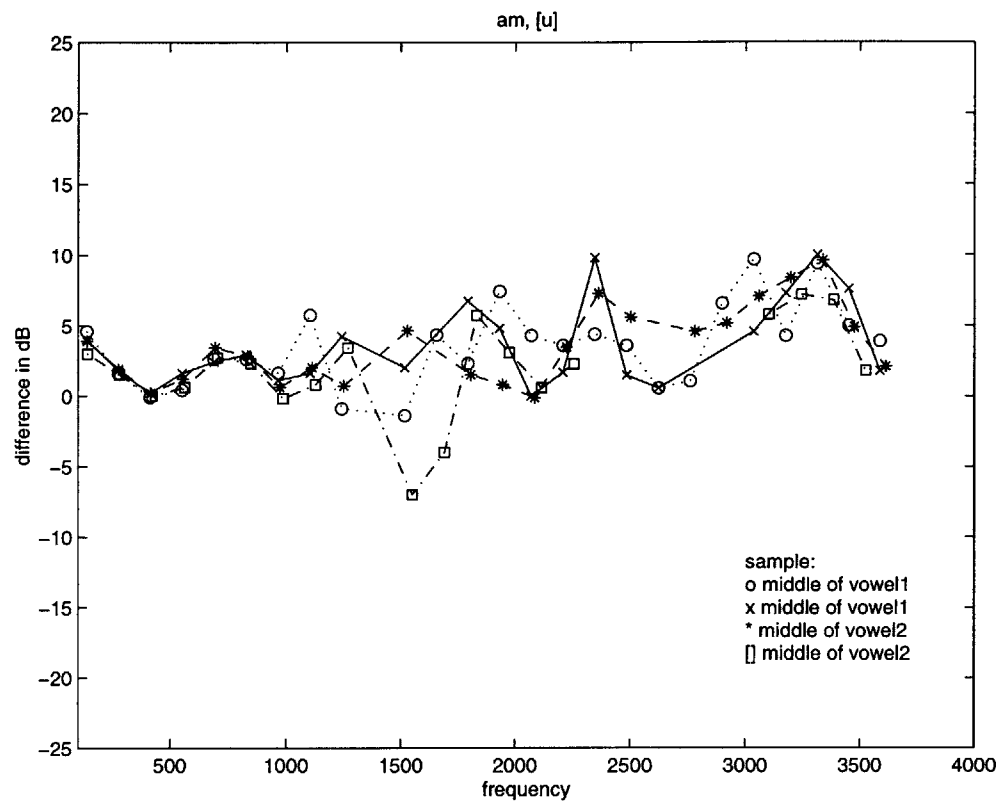
of the peak that follows almost immediately 200 Hz after it. It also appears that there is close tracking between the samples for those two spectra for which the data were reliable. Moreover, the general trend of the difference increasing with frequency can be ascertained from the plots.



**Figure 3.8:** [i] difference plot; comparison of all speakers

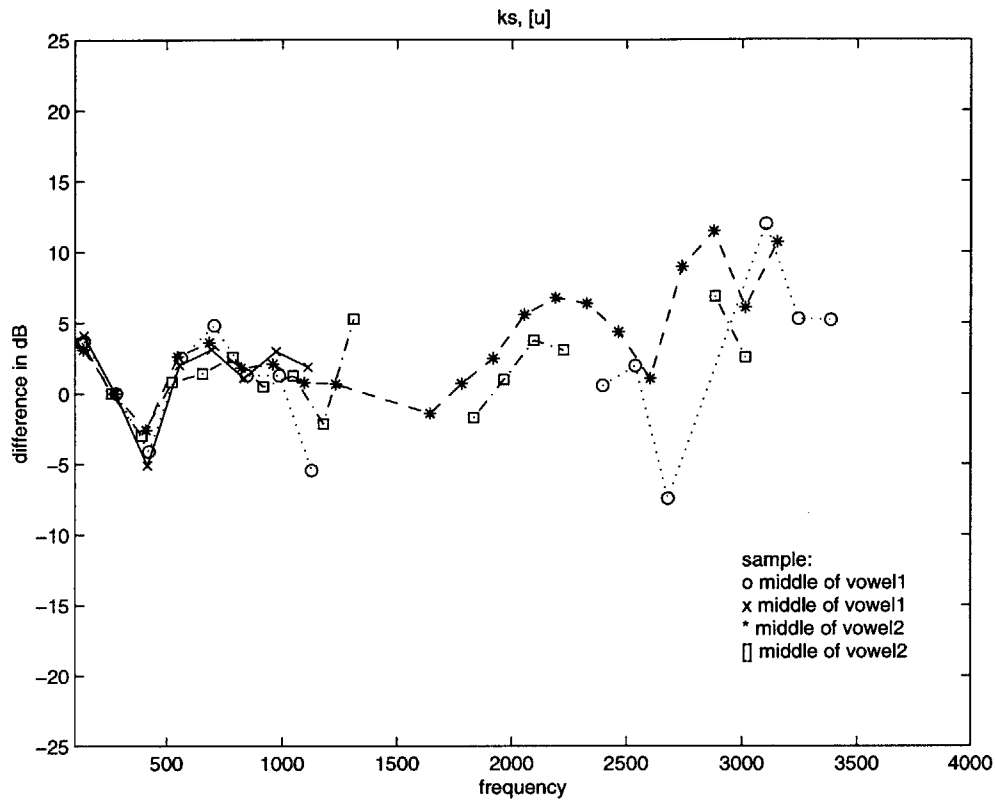
From the comparison of the difference plots (Figure 3.8) of each speaker it is easy to tell that the vowel [i] was not treated as equally by each speaker as was the vowel [a]. There is much more divergence between the difference plots of each speaker just above 500 Hz. This result is reliable because the samples of individual speakers tracked well with themselves in the lower and higher frequencies.

### 3.1.3 [u]



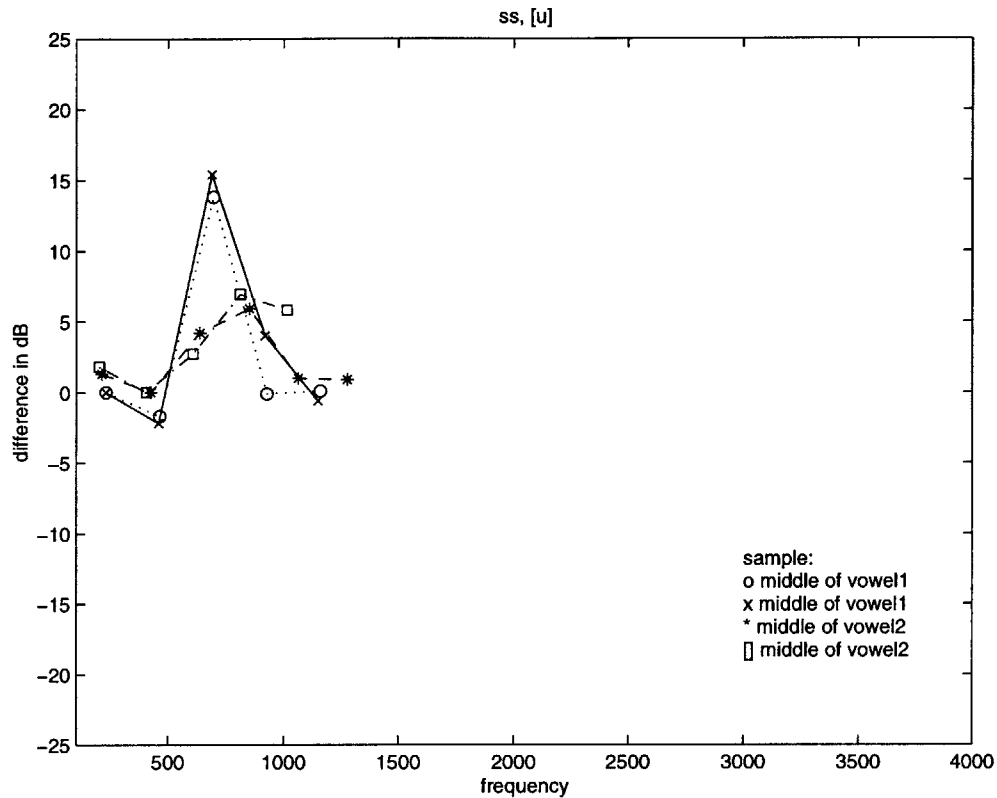
**Figure 3.9:** am [u] difference plot

AM's difference plots for the vowel [u] are pictured in Figure 3.9. Again, as with [a] and [i], the samples follow the same pattern and have nearly identical difference values for frequencies up to about 1000 Hz. For frequencies above 1000 Hz the samples diverge but seem to converge again after 3000 Hz. The 100 Hz to 1000 Hz frequency range of the difference plots resemble those of [a] in that the contour of the plot is smooth and after the first harmonic difference stays within a plus or minus 3 dB (absolute value) range. Once more there is an increasing positive trend in the differences with an increase in frequency.



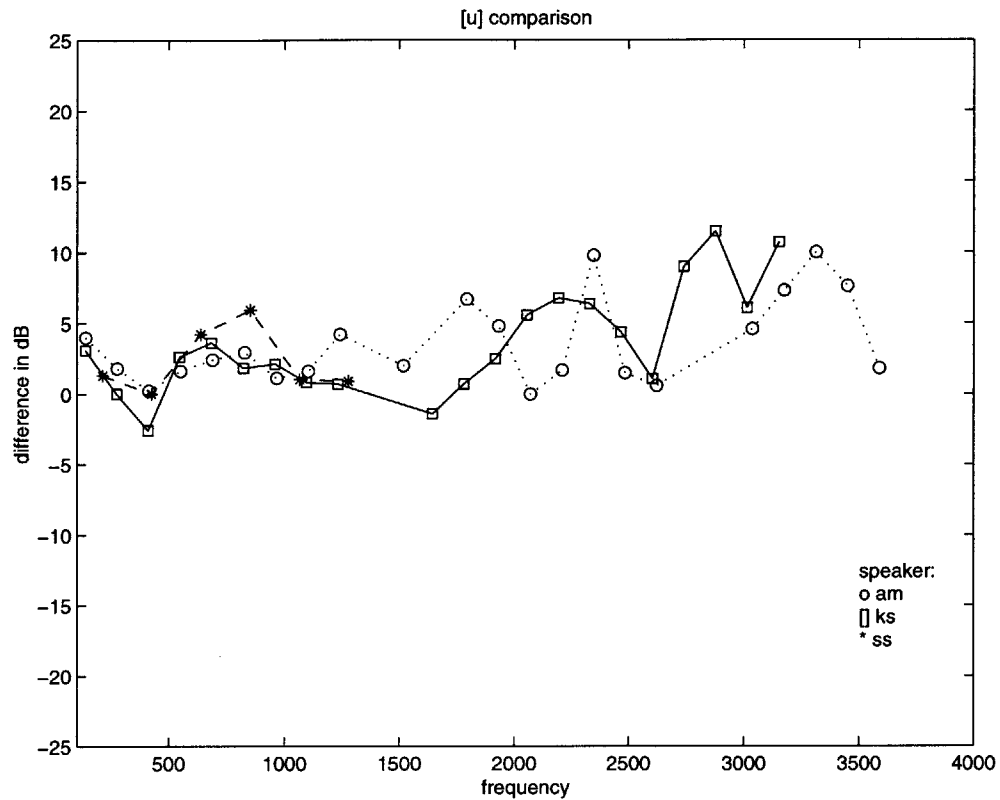
**Figure 3.10:** ks [u] difference plot

For speaker ks, the difference plot (Figure 3.10) actually resembles that of [i] more than [a]. There is a sharp decrease in difference around 450 Hz as there is in the [i] plot; the only difference is the absence of a major peak following a 200 Hz increase in frequency. There is a lack of data from 1000 Hz on due to noise. However, there is some correlation to be seen between the sparse data points of each sample's difference plot in the higher frequencies. Once again, there seems to be a general increase in the differences as frequency increases.



**Figure 3.11:** ss [u] difference plot

Speaker ss has an [u] difference plot (Figure 3.11) that has a pattern that is similar to her difference plot of [a], as does speaker am. In this situation there was a total loss of higher frequency information due to a lot of noise in the higher frequency region of the spectra. The vowel [u] has a weak third formant peak and therefore the signal-to-noise ratio is lesser in that region. However, from the data that is available it can be seen that there is close tracking in the pattern of the samples, and that, in this instance, ss has a large difference between the spectra below 1000 Hz. This peak in difference is about 15 dB, which is actually 5 dB greater than the peak in the [a] difference plot.



**Figure 3.12:** [u] difference plot; comparison of all speakers

Comparing the three speakers leads to the realization that something different is happening for each speaker in the way in which each microphone is transducing the vowel [u]. The comparison of the difference plots of each speaker can be seen in figure 3.12 above. The difference plots for the [u]'s of ss seem to have peaks around 750 or 800 Hz while those of am and ks have differences approaching zero.

### 3.2 Analysis

The same general pattern in the low frequencies indicates that the utterances of each speaker are being treated equally by both microphones. The fact that the difference measures for the first har-

monics for ss are less than those for ks and am is encouraging. The vocal pitch of ss is always higher and consequently the frequency of the first harmonic for ss always falls between the first and second harmonic frequencies of ks and am. Since the difference has decreased from the first harmonic to the second on each sample of each vowel it seems consistent that the first harmonic differences of ss should be less than those of ks and am since it is located in a frequency region where the pattern shows that the difference should be decreasing. The only major discrepancy is the difference plots of ks's [u] not looking like [a] as did those of am and ss. It actually seems more logical, however, that the difference plots of [i] and [a] would look more similar since both are pronounced with spread lips.

In the higher frequencies there was not much correlation between the difference plots in the interspeaker results, or between each speaker's own samples. The exception to this was am's [a] difference plot where the difference plot of each token was nearly congruent with others. However, am's other difference plots exhibit the same behavior in the higher frequencies as the other two speakers. One possible explanation for this is that during the [a], am's production of the sound was very steady, i.e. he had very little articulatory movements in the tongue. While slight shifts in the tongue will not perturb the lower frequency components of a sound they could easily cause shifts in the higher frequency components. Yet, another possibility (although it seems less likely due to am's [a] difference plot) is an issue of phase. Because of the location of each microphone, they may not be receiving the higher frequencies of the sound in the same relative phase. The windowing exaggerates this effect causing the difference plots to look different in the higher frequencies over each sample.



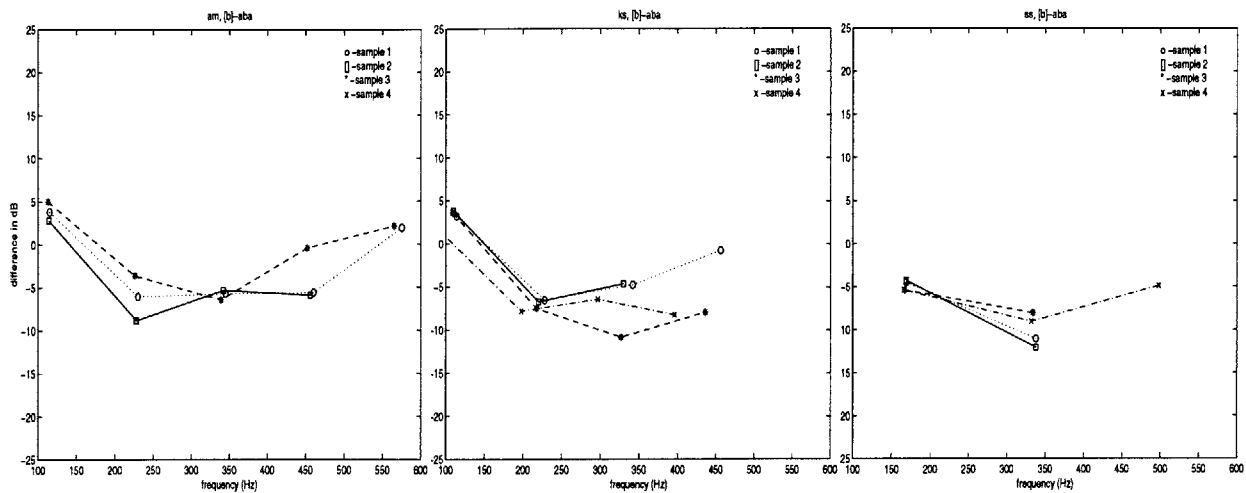
For the most part, the observed results follow the behavior that was predicted with the exception of the first harmonic difference. In the lower frequency range, where the sources were expected to look like simple sources, there was very little difference in the spectra, as expected. The disparity in the differences of the amplitude of the first harmonic in each microphone's spectra could possibly be a consequence of preamplifier response. It is not only possible that the first harmonic difference was being caused by the channel, but near field radiation characteristic as well. However, the channel argument seems more likely due to the fact that the three vowels studied have somewhat different radiation characteristics. The result that came as a surprise was the fact that the close-talking microphone seemed to have a stronger spectra in the higher frequency for all the vowels. However, this may be partially explained by the close-talking microphones frequency response (Figure 1.3) which appears to give the spectrum at frequencies above 2000 Hz up to a 4 dB boost. However, there appears to be another cause for the increase in frequencies because some of the peaks in the higher frequencies exceed 10 dB.

# 4 Voiced Stop Consonants

## 4.1 Results

As described earlier, all of the signals used for data have been normalized relative to the adjacent vowel in the utterances. All of the difference plots in this chapter have the following origin for their samples. The first sample was taken from spectra created by using a window at the beginning of the voice bar just after the vowel. At this point the voice bar is strongest and should exhibit some properties of the adjacent vowel. The second sample was usually taken from a portion of the voice bar such that the corresponding time window for its spectrum did not overlap (or barely overlapped) with the window of the spectrum used for the first sample. The third spectrum was taken in the same manner but with respect to the window used for the second sample. Some difference plots for these stop consonants may contain four samples, others may only have two, depending on the length of the voice bar and the amount of noise in the signal (because the voice bar gets weaker as it continues and consequently so does the signal-to-noise ratio). Some plots of samples were omitted because of noise. However, most of the difference plots contain three samples from each stop consonant.

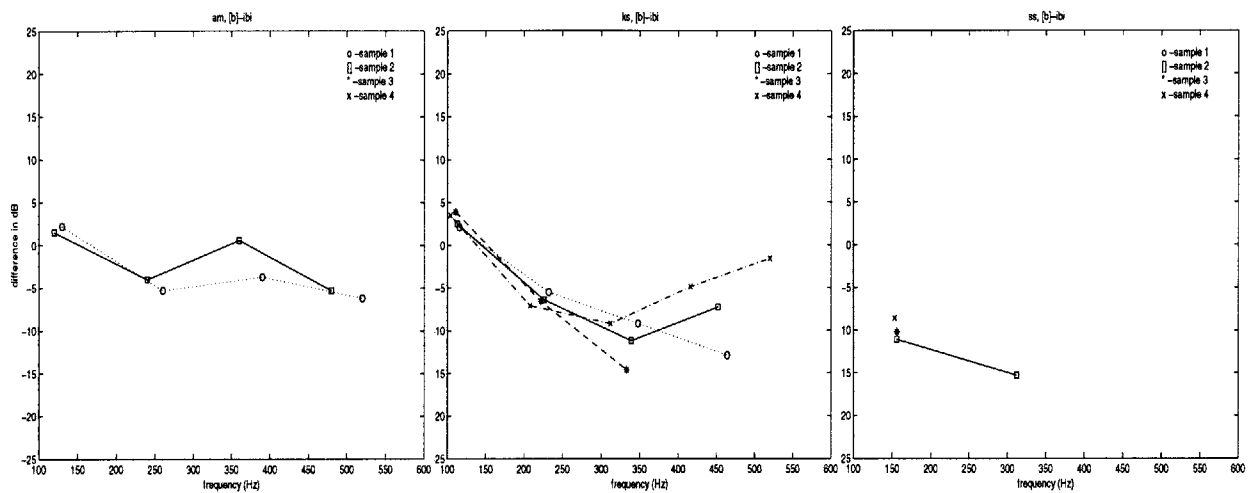
### 4.1.1 [b]



**Figure 4.1:** [b] from aba difference plots for all speakers

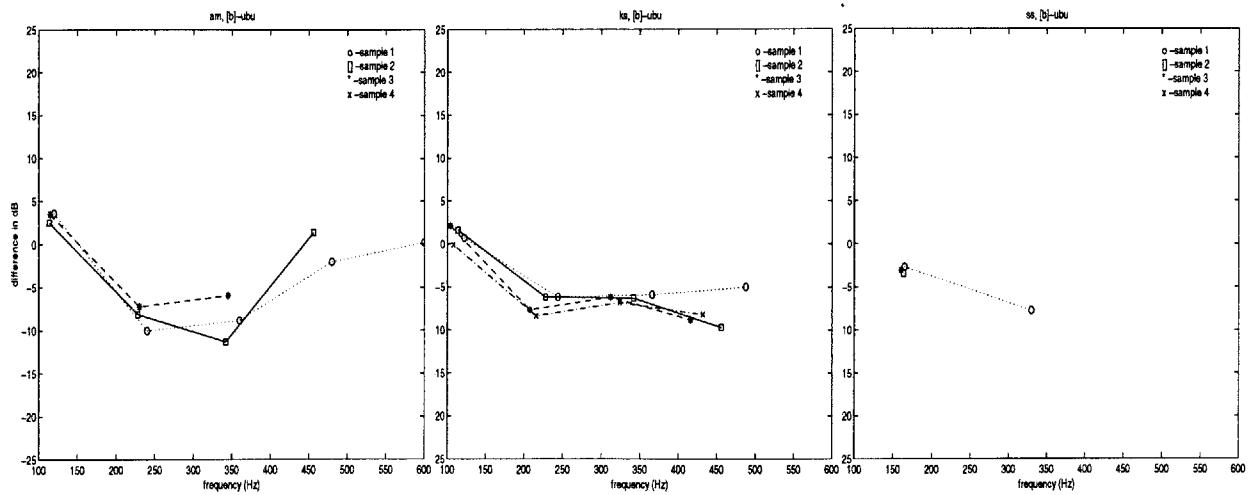
Figure 4.1 displays the difference plots of the spectra taken from the consonant [b] of the utterance aba for each speaker. The difference plots for am's [b] is pictured on the left. It can be seen that for each sample there is a negative 10 dB change in the differences of the first and second harmonics. In each instance the first harmonic was about 4 dB greater in the closetalking microphone, but the second harmonic was about 6 dB greater in the omnidirectional microphone. However, after the second harmonic difference the patterns of each sample begin to diverge. On the third sample the difference continues to decrease (become more negative) for the third harmonic difference whereas on the other two samples experience an increase, although the differences are just about equal. The discrepancy in pattern continues for the fourth harmonic where the first and second samples converge and again the third diverges. For ks the same behavior in the first two harmonics is seen in all four samples. The first second and fourth samples continue to follow the same pattern of difference through the third harmonic difference but the third sample does not. However, it does share a similar characteristic of am's plot in which the third harmonic's absolute difference is less than that of the second harmonic. For speaker ss the signal-to-

noise ratio was low so the amplitude of the harmonics beyond the first two were unreliable. Therefore only the differences between the first and second harmonics are used. The first harmonic differences of each sample are consistently about -5 dB. The first and second samples then experience an increase in the absolute difference of the second harmonic to about -13 dB, while the absolute difference only increases to about -9 dB on the second two samples



**Figure 4.2:** [b] from ibi difference plots for all speakers

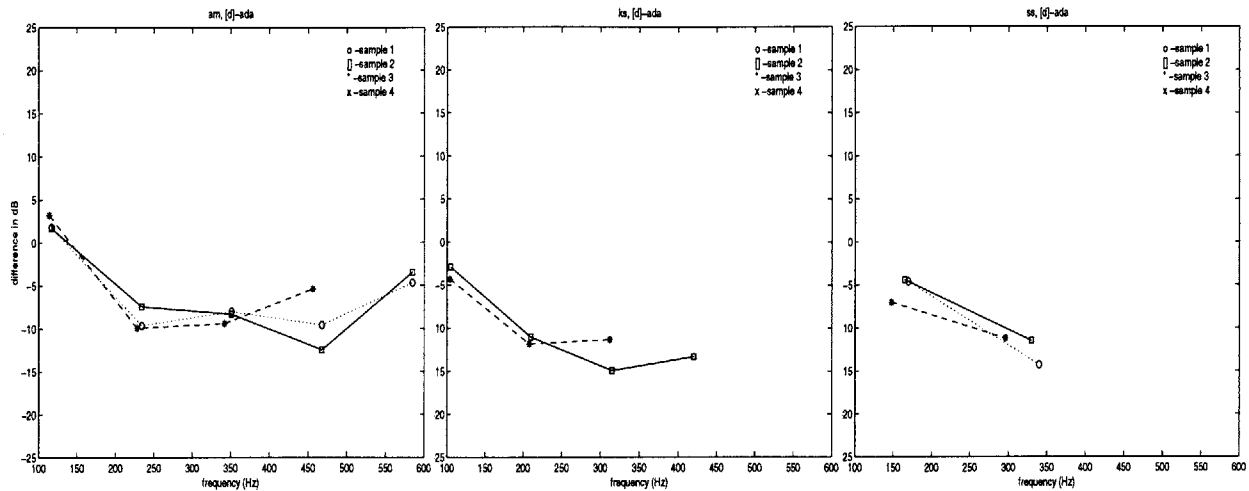
The similarity in results do not continue for the [b] plots made from the spectra of the utterance ibi (plots shown in Figure 4.2). Again, for both am and ks the first harmonic difference is positive and the second harmonic difference is negative. For ss the first harmonic difference is negative. Those are the only similarities to the [b] difference plots of aba. In am's difference plot the absolute difference decreases for the third harmonic only to increase again for the fourth. The difference plots of ks display different behavior in that on each sample the absolute difference increases for the third harmonic. The results for the fourth harmonic are varied across samples. For ss the spectra were too noisy to extract accurate values of the harmonic amplitudes.



**Figure 4.3:** [b] from ubu difference plots for all speakers

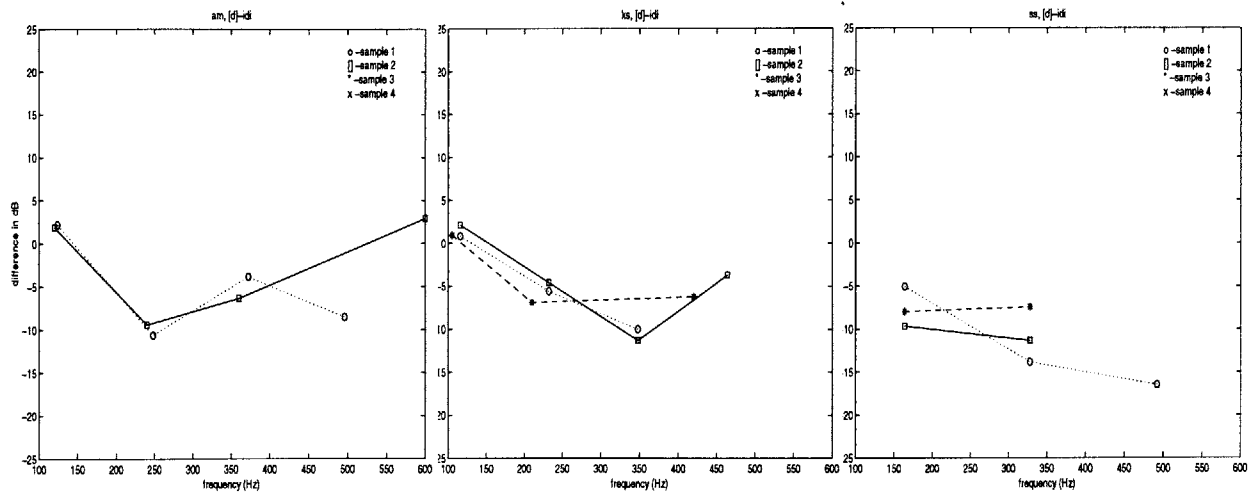
The dissimilarities between speakers also holds true for the difference plots of [b] created from the spectra of the utterance ubu. These plots are pictured in Figure 4.3. There is the same -10 dB change in the difference between the first and the second harmonic differences in am's difference plot as there was for the previous two. However, the samples diverge in pattern for the results for the third harmonic and above. For the plots of ks there is consistency in the pattern of the samples with the exception of the fourth harmonic difference of the first sample. Again there is consistency in the change in difference between the first and second harmonic differences. There is not much information to be acquired from the difference plots of ss, again due to noise. But the consistency of the differences between the first harmonics across all samples should be noted.

### 4.1.2 [d]



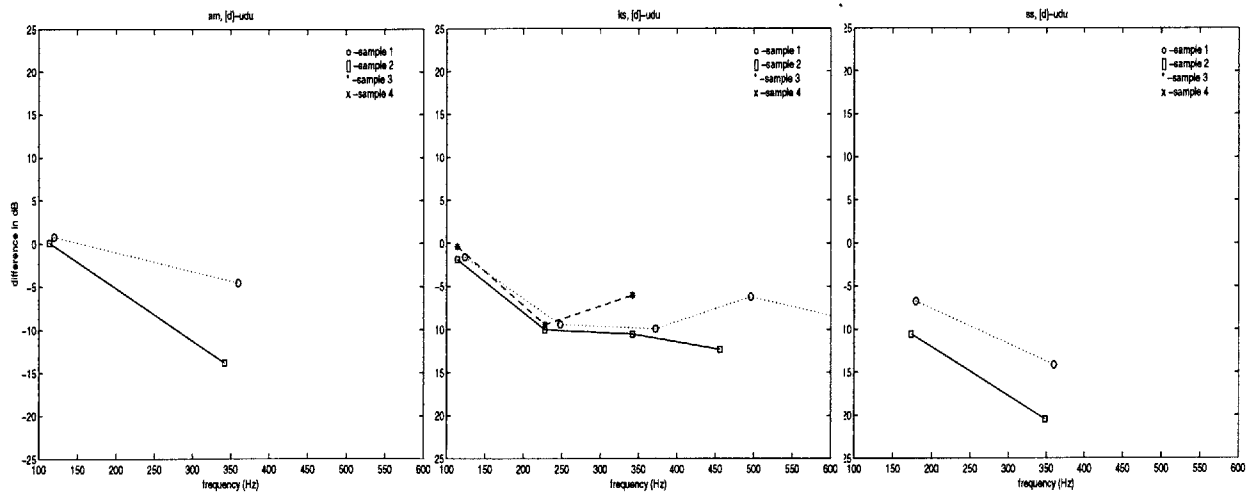
**Figure 4.4:** [d] from ada difference plots for all speakers

Many of the same trends observed in the [b] difference plots appeared in the results for the [d]'s. Figure 4.4 displays the difference plots for the [d]'s from the utterance ada. As was the case for his [b] plots, am's plots reveal an 8 to 10 dB decrease in the differences between the first and second harmonic differences. The absolute difference between the third harmonics decreases slightly from that of the set of second harmonics of every sample. However, the results of the set of differences between the fourth harmonics are varied. For ks the first harmonic difference is negative. The second harmonics' absolute difference increases another 7 dB or so, continuing with the pattern observed in the [b] difference plots. Because of the disparity in data no conclusion can be drawn for the behavior of the third harmonic difference. More of the same first and second harmonic differences can be seen in the [d] plots for ss.



**Figure 4.5:** [d] from idi difference plots for all speakers

Similar behavior is witnessed for the results from the [d] difference plots (in Figure 4.5) made from the utterance idi. The disparity in the first and second harmonic differences can be seen for all three speakers. The largest disparity again occurring in the plots of am. However, the first harmonic difference of ks is again positive. The absolute difference of the third harmonic increases for two of ks's samples while it decreases in am's samples. There is so much variance in the samples of ss's difference plots it is difficult to tell which is reliable, if any at all.

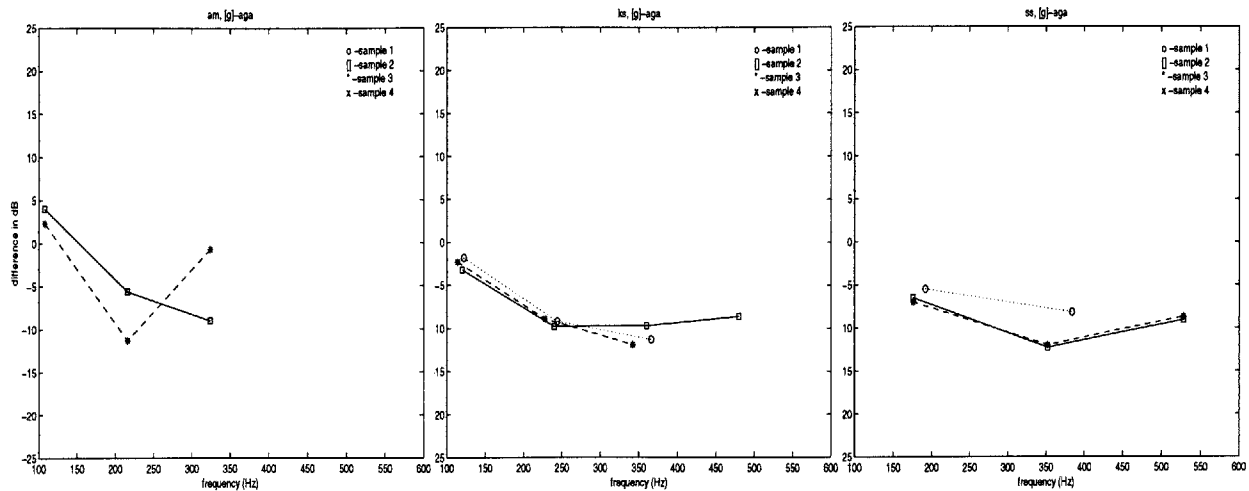


**Figure 4.6:** [d] from udu difference plots for all speakers

The difference plots of [d] (in Figure 4.6) from the utterance udu do not depart much from the previous observations. In this case, however, the difference plots of am do not yield any solid information because of the large disparity in the second harmonic difference of the two samples. The plots of ks reveal a similar first harmonic to second harmonic difference change to the previous two [d] plots, but in this case the difference of the first harmonic is close to zero. In two of the samples the third harmonic difference does not vary much from the second. For ss the first and second harmonic differences are different on each sample but the change between the two is about the same in both.



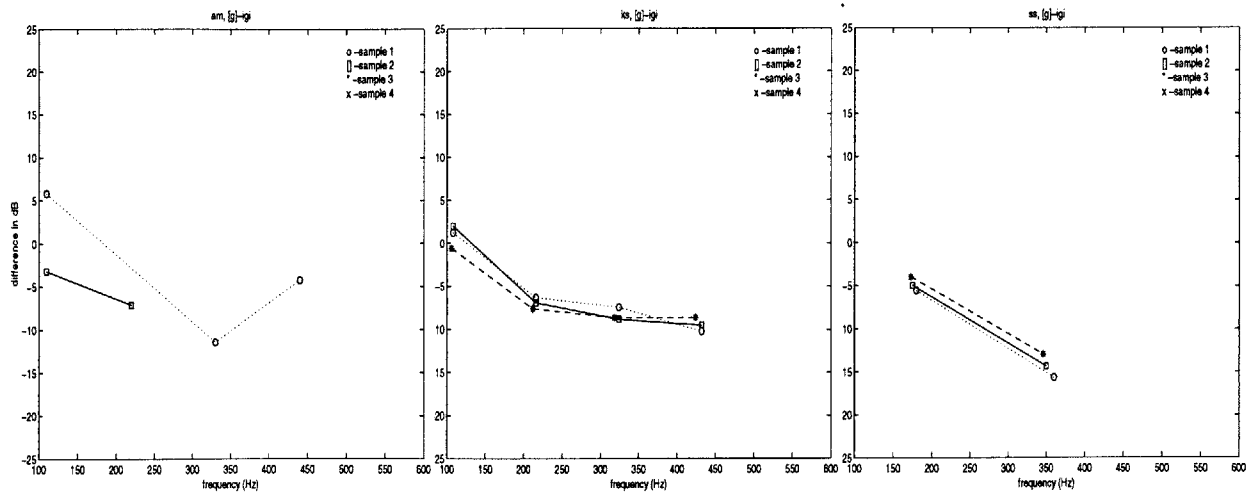
### 4.1.3 [g]



**Figure 4.7:** [g] from aga difference plots for all speakers

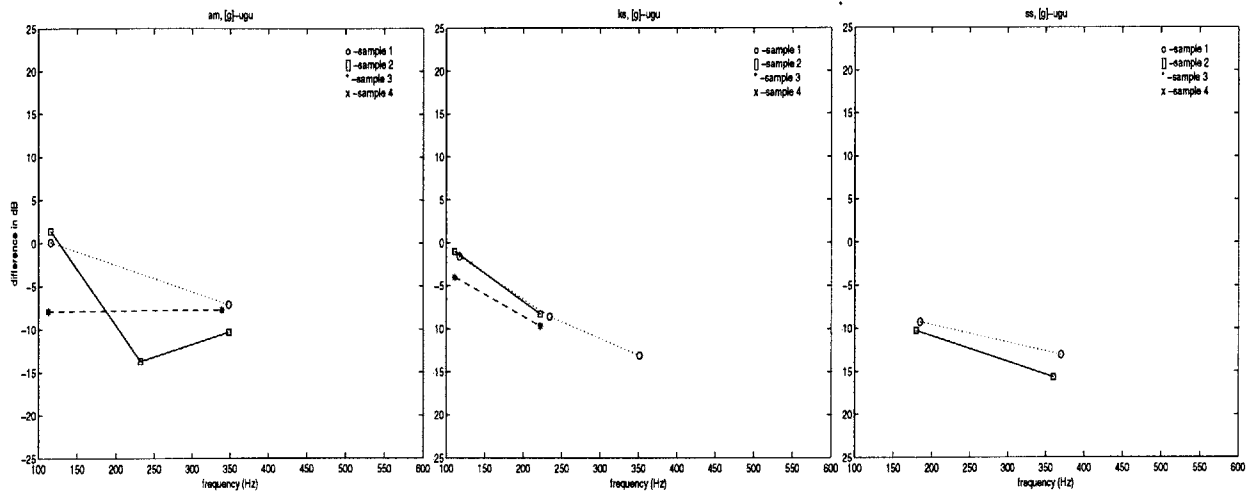
Figure 4.7 contains the difference plots for [g] for all speakers derived from the utterance aga.

Starting again with am, the change from the first to second harmonic difference is about 10 to 12 dB on samples two and three respectively. The change in the first to second harmonic difference is not as large for ks, only about 7 on each sample. The first harmonic difference of the am plots are positive and in those of ks they are negative. The third harmonic difference does not change much from the second. For ss the change in first to second harmonic difference is even less, only about 6 dB on the second two samples and about 4 dB on the first. The first harmonic difference is negative as it is for ks. The absolute difference of the third harmonic decreases about 4 dB on the second and third samples.



**Figure 4.8:** [g] from igi difference plots for all speakers

In figure 4.8, which shows the difference plots for [g] imbedded in the vowel [i], it can be seen that the plots for am show two very different results. One sample has a positive first harmonic difference and the other has a negative first harmonic difference. Because, there is no second harmonic difference in the first sample not much else can be drawn from the plots. All of the samples in ks's plot are fairly consistent, with the first harmonic difference beginning between 0 and 2 dB and the second harmonic difference decreasing to about -8 dB. The two following harmonic differences remain within a few decibels of the second. All of ss's plots are consistent as well. However, they do not seem consistent with the plots of ks. The first harmonic difference of ss is located in the frequency range of the second harmonic difference of ks, but its value is a bit higher. But the difference is really in the second harmonic difference of ss which drops about 10 dB down from the first. In the ks plot the harmonic differences in this region have only decreased by 2 to 3dB from the second harmonic difference.



**Figure 4.9:** [g] from ugu difference plots for all speakers

Figure 4.9 contains the difference plots for the [g] spectra created from the utterance ugu. Again, am's samples do not appear consistent with one another. On the first sample the first harmonic difference is positive. This is also true for the second sample but not the third. Only the second sample plot has a second harmonic difference so the behavior in that region can not be ascertained. However, the values of the third harmonic difference on each sample are not too spread out in value. The plots of ks have more consistency. The values of the first and second samples almost overlap, and although the same is not true for the third sample plot it is parallel to the other two sample plots. Each first to second harmonic difference decreases between 7 to 10 dB, unlike the plots of am. The plots of ss are somewhat consistent as well and almost appear to follow the same pattern in frequency as ks.

#### 4.1.4 Comparison of each Voiced Stop Consonant across the vowels

**Table 4.1:** [b] 1st and 2nd harmonic differences across vowels

Vowel Context	am	ks	ss
[a]	1st: 3.9 2nd: -6.1 Diff: 10	1st: 2.9 2nd: -7.1 Diff: 10	1st: -4.9 2nd: -10.1 Diff: 5.2
[i]	1st: 2.0 2nd: -6.1 Diff: 8.1	1st: 3.0 2nd: -6.4 Diff: 9.4	1st: -10 2nd: N/A Diff: N/A
[u]	1st: 3.2 2nd: -8.4 Diff: 11.6	1st: 1.1 2nd: -7.1 Diff: 8.2	1st: -3.1 2nd: N/A Diff: N/A

Table 4.1 contains the data of the averages of the first and second harmonic differences as well as the difference between the first and second differences for [b] across all vowels. The table was created to compare how each [b] spectral difference changed with vowel context. Although it is not expected that individuals will have the same value differences it is expected that the relative change in the difference while changing vowel context would be about the same. The first two harmonic differences were chosen for display because they were fairly consistent throughout all samples. The difference between the two is displayed because, at least for ks and am, they have similar fundamental frequencies, it is expected the differences between the first two harmonics should change proportional for each speaker from one vowel context to the next. However, none of the expected behavior is witnessed in the table. The smallest first harmonic difference average for am is during the vowel [i] and this vowel context has the greatest first harmonic difference for ss. The smallest first harmonic difference occurs for ks's [b] in the context of [u]. The second harmonic differences between ks and am are not far apart. The second harmonic difference data for ss in the context of [i] and [u] were not available due to noise.

**Table 4.2:** [d] 1st and 2nd harmonic differences across vowels

Vowel Context	am	ks	ss
[a]	1st: 2.2 2nd: -9.0 Diff: 11.2	1st: -3.6 2nd: -11.4 Diff: 7.8	1st: -5.3 2nd: -12.3 Diff: 7
[i]	1st: 2.1 2nd: -10 Diff: 12.1	1st: 1.5 2nd: -5.2 Diff: 6.7	1st: -7.6 2nd: -10.9 Diff: 3.3
[u]	1st: N/A 2nd: N/A Diff: N/A	1st: -1.3 2nd: -9.6 Diff: 8.3	1st: -9.7 2nd: -18.1 Diff: 8.4

The table format of 4.2 is the same as that of 4.1. The purpose of this table is to compare the [d] closure vowels across vowel context for the same reasons given in the previous paragraph. However, in this case, the first harmonic and second harmonic differences are not expected to vary much as a consequence of vowel context because the articulatory tongue movement of [d] is not affected much by vowel context. Nevertheless there are still differences in the vocal tract shape. There is not much variation in the am data, although the [u] data was too noisy to use. However the other two speakers show variations as large as 5 dB in the first harmonic difference and about 8 in the second.

**Table 4.3:** [g] 1st and 2nd harmonic differences across vowels

Vowel Context	am	ks	ss
[a]	1st: 3.2 2nd: -8.5 Diff: 11.7	1st: -2.4 2nd: -9.7 Diff: 7.3	1st: -6.7 2nd: -10.8 Diff: 4.1
[i]	1st: 3.2 2nd: -7.1 Diff: 10.3	1st: 0.9 2nd: -6.9 Diff: 7.8	1st: -4.9 2nd: -14.4 Diff: 9.5
[u]	1st: 1.4 2nd: -13.7 Diff: 15.1	1st: -2.2 2nd: -8.9 Diff: 6.7	1st: -9.8 2nd: -14.6 Diff: 4.8

Table 4.3 displays the first and second harmonic difference data for the [g] closure across all vowel for all speakers. Again, as with [d], and for the same reason, there is not much variation expected as a result of vowel context. While there is some consistency in the data for each speaker there seems to be none between speakers. An example of this is that am's two similar vowel context (similar in values recorded) are [a] and [i] but for ks they are [a] and [u].

#### 4.1.5 Comparison of the Voiced Stop Consonants in the context of [a]

**Table 4.4:** Voiced Stop Consonant measurements in the vowel context of [a]

Consonant	am	ks	ss
[b]	1st: 3.9 2nd: -6.1 Diff: 10	1st: 2.9 2nd: -7.1 Diff: 10	1st: -4.9 2nd: -10.1 Diff: 5.2
[d]	1st: 2.2 2nd: -9.0 Diff: 11.2	1st: -3.6 2nd: -11.4 Diff: 7.8	1st: -3.6 2nd: -11.4 Diff: 7.8
[g]	1st: 3.2 2nd: -8.5 Diff: 11.7	1st: -2.4 2nd: -9.7 Diff: 7.3	1st: -6.7 2nd: -10.8 Diff: 4.1

Table 4.4 is read in the same manner as the previous tables: the columns contain the data of each speaker and each of the rows contains the information of one stop consonant for all speakers.

Each row contains the first and second harmonic difference information for that speaker for that consonant. The vowel [a] was chosen as an example because the other vowels show similar “randomness” in their results, but mainly because it is the only vowel for which all speakers had a full amount of data. Both am’s and ks’s average first harmonic differences are positive for the consonant [b]. Both first harmonic difference averages decrease for the consonant [d] as might be expected because of the smaller oral cavity during closure. However, for ks the difference becomes negative and the net change is 6.5 dB whereas for am the net decrease is only 1.7 dB. The difference actually increases positively for both am and ks from [d] to [g], but that means a smaller absolute difference in the case of ks and a larger absolute difference for am. For ss there is also a larger absolute difference with the average first harmonic difference increasing negatively from the [d] to [g]. There is no noticeable trend between the speakers data in the overall comparison of the voiced stop consonants within this context.

## 4.2 Analysis

Overall, there was too much variance between each speaker's results to understand exactly what was occurring. However, there are a couple of trends in the data common to each speaker. The first trend is each speaker having somewhat consistent results for the differences between the first harmonic and second harmonic differences. The change in first harmonic to second harmonic difference was about 10 dB on average for am and about 7.5 dB on average for ks. The results for ss were scattered. Secondly, on average, the first harmonic differences were higher for b, in this case meaning more positive, than those of d and g for all speakers. This result is not very surprising. The result that was most surprising was the fact that the first harmonic had a positive difference even for the consonant [g]. It is expected that most sound during the [g] voicebar would radiate entirely from the neck surface and thus show a weaker amplitude for the close-talking microphone. However, the results show that this may not be the case. Because, each speaker pronounced the utterances in the same manner (no accents, etc.) it is likely that the large variance in results can be attributed to differences in anatomy. That is, the size differences of the vocal tract are not scalar or necessarily proportional. For example, males generally have longer vocal tracts than females. That does not mean, however, that for a particular vocal tract configuration the back to front cavity length ratio will be the same for male and female speakers. A female may have a large front cavity and small back cavity while a male may have a back cavity length proportional to that of his front cavity of his vocal tract. Such differences will lead to different resonance prominences, such that the front may dominate the back or the back may dominate the front at certain frequencies, leading to the disparate results between speakers as is witnessed in the results of this section, though such large differences are not expected at low frequencies.

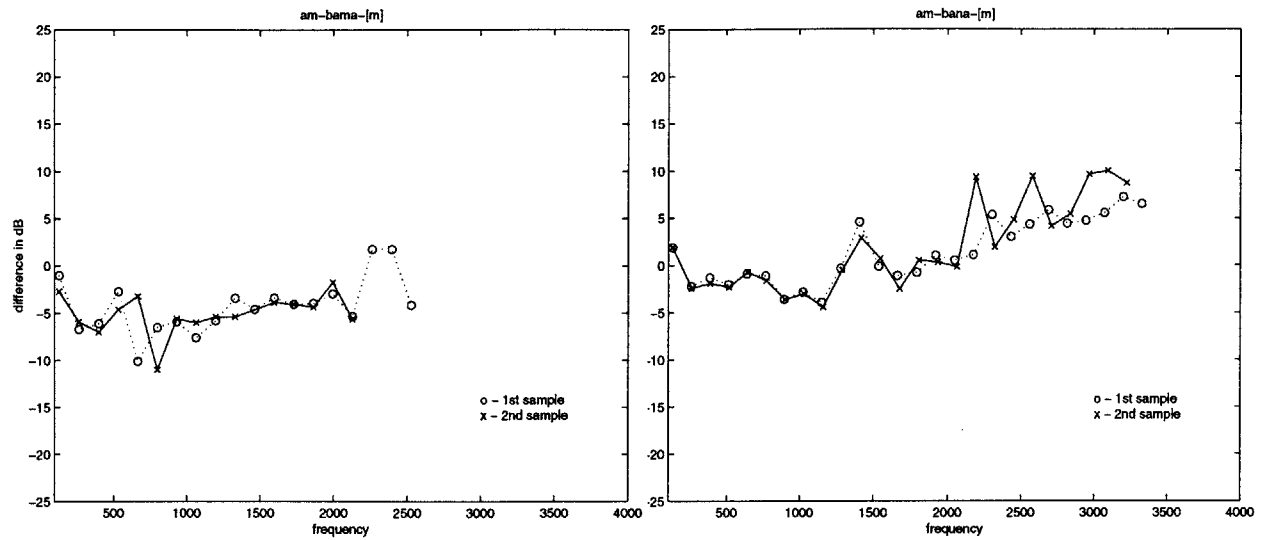


# 5 Nasal Consonants

## 5.1 Results

The nasals were studied in the same manner as the stop consonants. Each of the utterances containing the nasals [m] and [n] also contained one of the vowels [a], [i], or [u] and began with the stop consonant [b]. The [b] was placed at the beginning of each of these utterances so that the initial part of the vowel which followed it would not be nasalized. These un-nasalized portions of the vowels were used to normalize the spectra of the two microphones such that the peak of the first formant was equal. Once the signals were normalized, two spectra of the nasal part of the utterances were created by windowing a section of the signal after the first few pitch periods of the nasal murmur and windowing another section before the last few pitch periods. These sections were selected in order to get two separate (no overlap) samples of spectra taken at steady portions of the nasals. Difference plots were made from these spectra and are presented in the following sections. The results are presented by utterance. They are presented in this manner in order to examine [m] and [n] in the same phonetic context since they were predicted to have similar differences. When [m] and [n] are in the same vowel context they should have the same low frequency resonance, and therefore any radiation from the neck surface may be similar. The sound radiating from the nose may be somewhat similar but the zero in the spectra should be located at a different frequency for each due to the different articulations of [m] and [n].

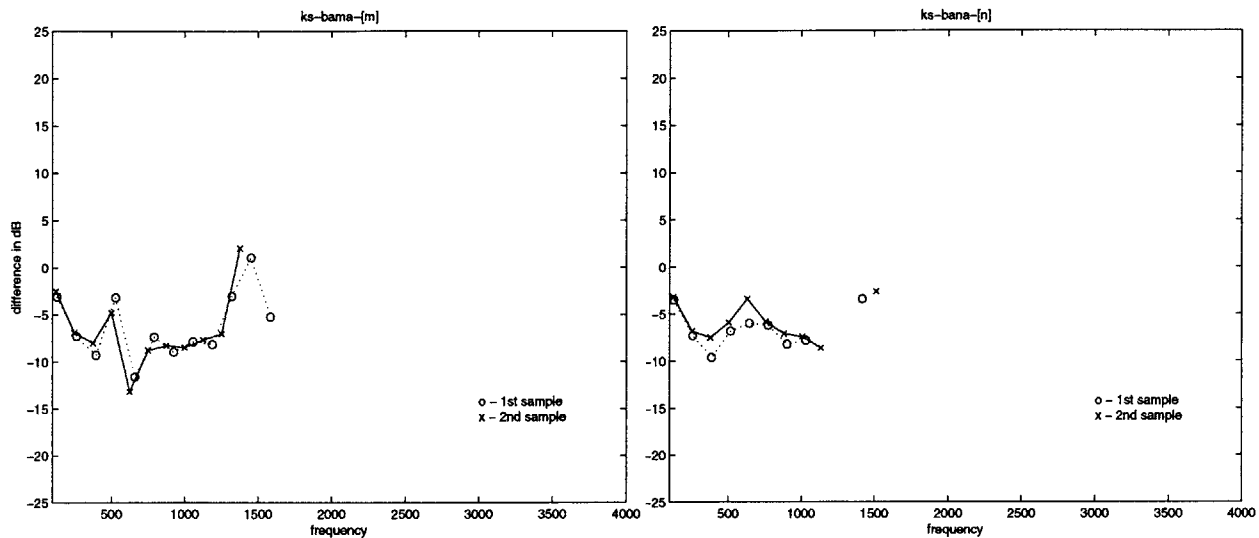
### 5.1.1 [m] and [n] from bama and bana



**Figure 5.1:** am [m] and [n] difference plots for bama and bana

Figure 5.1 above displays the difference plots of the [m] (right) and [n] spectra taken from the utterances bama and bana of speaker am. The difference plots of the [m] show a minima of about -7 or -8 dB between 300 and 500 Hz on both samples. The difference increases by a few decibels for the next two harmonics and then experiences a sharp, negative peak at about 600 or 800 Hz, depending on the sample. Then the difference plot levels out to about -5 dB at 1000 Hz up to about 1900 Hz. Both sample's plots are fairly congruent in this region with very little discrepancy. The difference plots for the [n] of ana show no similar characteristics to that of [m]. One example of this disparity that is obvious is the fact that the first harmonic difference is positive while in the [m] plot it is negative. Also, the points in the difference plots never attain values lower than -5 dB, in addition to the fact that all of its peaks are in the positive direction. The [n] plot never levels out to a stable (not more than a one decibel variation) set of values except for the region which contains the second through fifth harmonic differences. As was the case with the [m] plots, the plots of both samples have the same shape and nearly equal values up to 2000 Hz

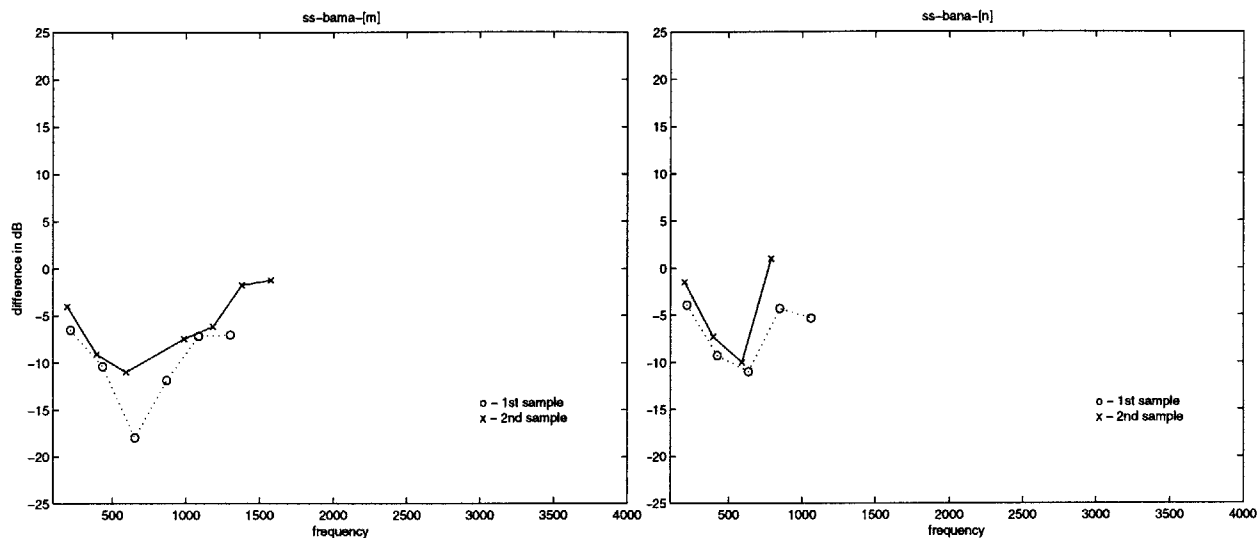
where they begin to diverge, so that the information provided is probably an accurate depiction of the spectral differences.



**Figure 5.2:** ks [m] and [n] difference plots for bama and bana

The difference plots for ks's [m] and [n] from bama and bana are given in Figure 5.2. Due to a lot of noise in the signal of the omnidirectional microphone in both utterances no data above about 1500 Hz could be used due to uncertainty in its accuracy of the harmonic amplitudes. A similarity between the [m] plots of am and the ks [m] plots by can be seen by looking at the data that was obtained for the ks [m] plots. There is a minimum of about -8 dB and -10 dB difference at the third harmonic on each sample of both plots and that is followed by a larger drop only 2 harmonics later. The plots level out to about -9 dB at 700 to 1100 Hz after which there is a peak up to 2 dB difference. But unlike the plots of am, the ks [n] difference plots are somewhat similar to the [m] difference plots. The first harmonic difference starts at approximately the same value and the next few differences share similar trajectory and value, through about 400 Hz, but the similarity ends there. The trajectory in the difference plot continues upward toward positive difference

although it is never rises above -3 dB at which point it slowly declines back down to its lowest point, but this is not at all congruent with what can be seen in the difference plots of [m].

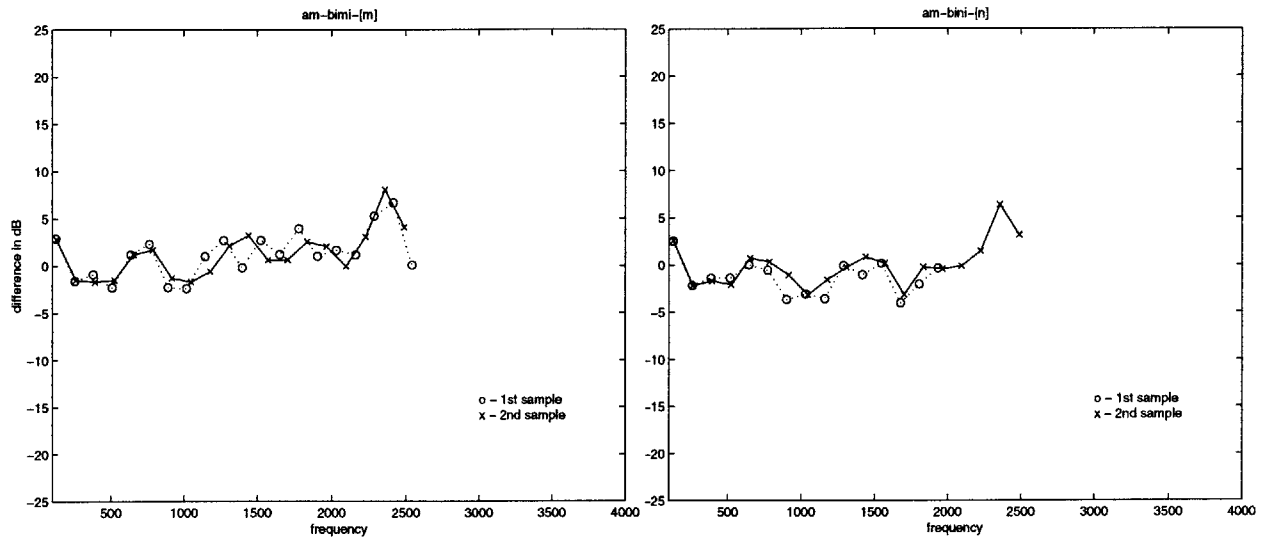


**Figure 5.3:** ss [m] and [n] difference plots for bama and bana

Figure 5.3 contains the difference plots for the [m] and [n] of ss from bama and bana. They follow the example of the difference plots of am and ks in that the plots of the two nasals appear to share only a little similarity. There is a bit of a sparse data problem for the plots of ss because of noise in the higher frequencies and simply because she has a higher fundamental frequency, leading to fewer harmonics. In the [m] plots the first harmonic difference is at -4 and -6 dB whereas in the [n] plots they are -1 and -4dB. In the [n] plots both samples share similar shapes while this is not the case for the [m] plot where the third harmonic has a -19 dB difference on the first sample and only a -11 dB on the second. The only similarity discernible is the fact that the difference of the first three harmonics decreases, with more of decrease between the first and second than the second and third.

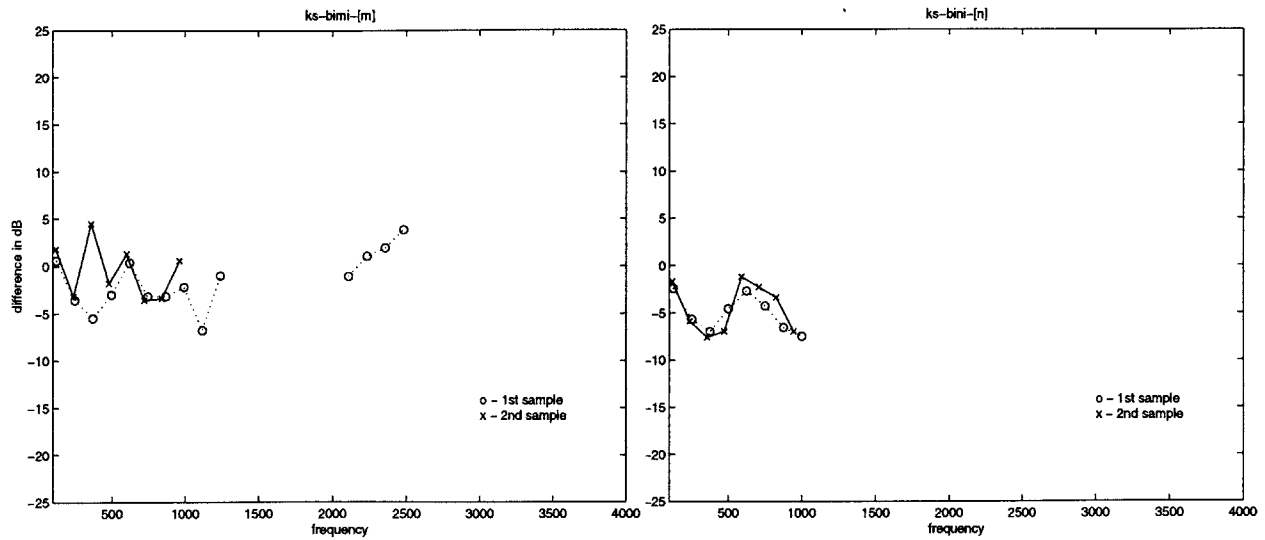
As alluded to in the previous paragraphs, there were some similarities between the [m] difference plots of am and ks. The shape made by the differences of the first five harmonics of am's plot is very similar, especially for the first sample of am, to both of those in the [m] plots of ks. The width of the shape of these differences of the first five harmonics is a little different between speakers because of the different fundamental frequencies of each speaker. In addition, the first three harmonics of am's [m] plot are similar to those of ks's [n] plot by associativity. But their [n] difference plots look nothing alike. A possible explanation for why ss's plots do not share any similarities with the those of the other two speakers could be that there are not enough harmonics in the difference plots of ss. For example, the large difference peaks in ss's plots that occur on the third sample may actually be due to the same phenomenon that causes the negative on the fifth sample of both male speakers. However, it is difficult to make such a judgment from so little information.

### 5.1.2 [m] and [n] from bimi and bini



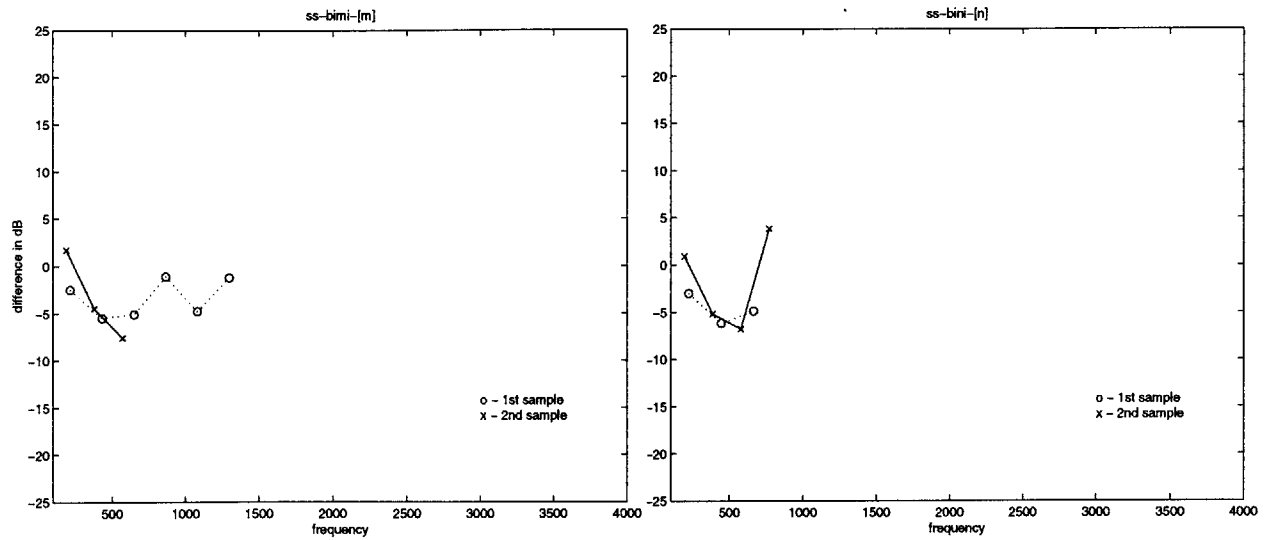
**Figure 5.4:** am [m] and [n] difference plots for bimi and bini

This section also begins with the discussion of the [m] and [n] plots of am which are given in Figure 5.4. The first harmonic difference of the [m] difference plots is about 3 dB on both samples and on the next harmonic the value in difference decreases to about -2 dB where it remains until a little over 500 Hz (the next three harmonics). The plots then increase in value to become slightly positive and then decrease to barely a (negative) difference at approximately 1100 Hz. At this point the plots of each sample begin to diverge. The same pattern is visible in the [n] difference plots of both samples through their first seven harmonic differences, at which point, the similarities between [m] and [n] diverge. However, unlike the [m] difference plots, the plots of [n] track fairly well throughout. Another similarity is found in the higher frequencies where on both second samples of the [m] and [n] difference plots is a 9 dB peak which occurs at about 2300 Hz.



**Figure 5.5:** ks [m] and [n] difference plots for bimi and bini

The difference plots of ks for [m] and [n] in this context do not appear to share any similarity with those of am (Figure 5.5). However, there is some similarity between themselves. The similarity is found in the first five harmonic differences of each plot. In both plots the third harmonic is a negative peak although the value is a couple decibels lower in the [n] plot. Moreover, in the [m] plot the second sample's third harmonic difference is actually a peak in the opposite direction but this could be anomalous, since the pattern of the first sample seems to match the pattern of both samples of the [n] plot so well. The main difference between the two plots in this region between 100 and 700 Hz is the fact that the first and fifth harmonic differences in the [m] plot are slightly positive whereas in the [n] plot they are both about -2 dB. After the fifth harmonic difference the trajectories of the plots begin to diverge but it is not clear how long this trend would continue, again because of a lack of data due to noise.

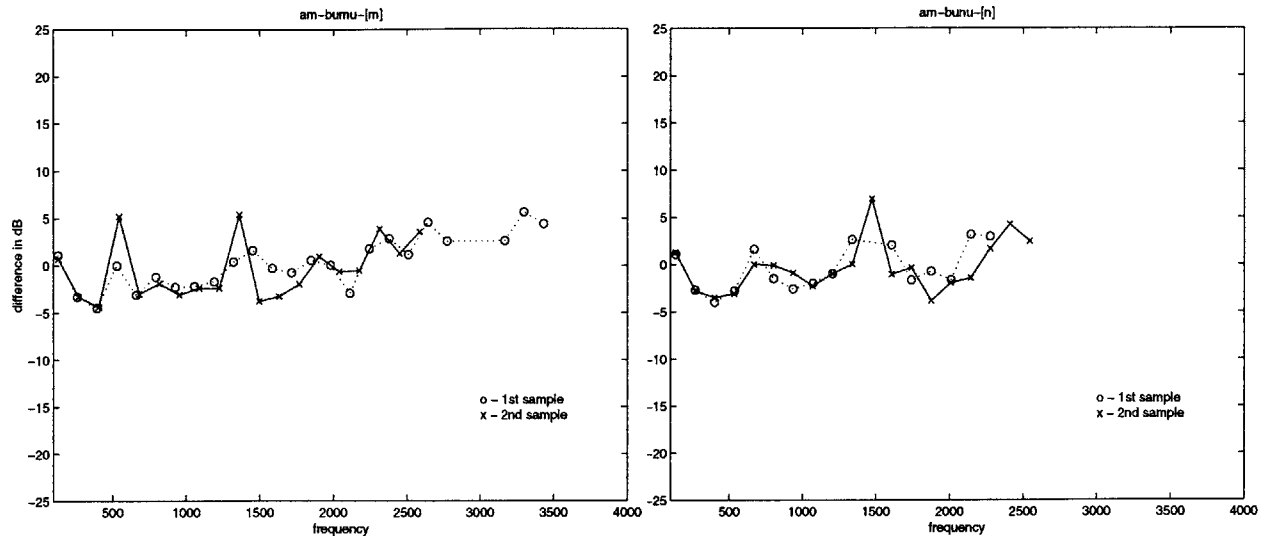


**Figure 5.6:** ss [m] and [n] difference plots for bimi and bini

The number of data points in the [m] and [n] difference plots of ss (Figure 5.6) were also limited due to noise. However, from the information that is attainable it can be seen that the first three harmonic differences of both the [m] and [n] difference plots are similar for the first samples as well as the second pair of samples in shape and in value, although in each the set of samples have two different patterns. This could have been due to the way that ss's nasal transitioned into the following vowel, changing the characteristic of the spectra and thus the differences. This would show up in the data if the sample was taken at a transitory point in the signal instead of a more steady portion. The first four points of the first sample in the [m] plots looks as if it could be following a pattern similar to the first six points of the [m] and [n] plots of am.

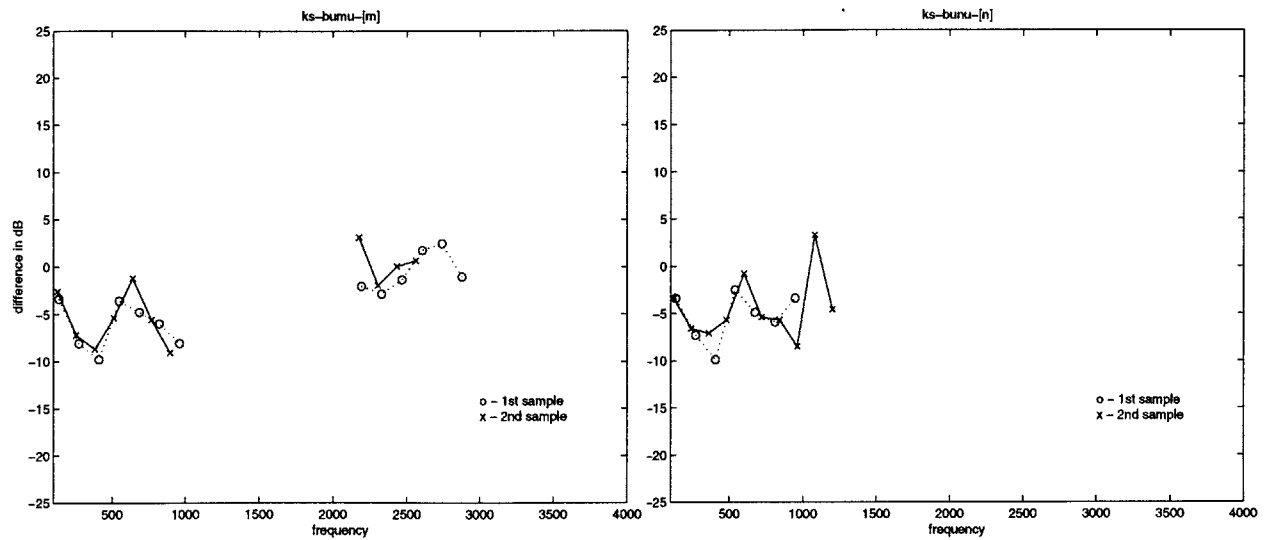


### 5.1.3 [m] and [n] from bumu and bunu



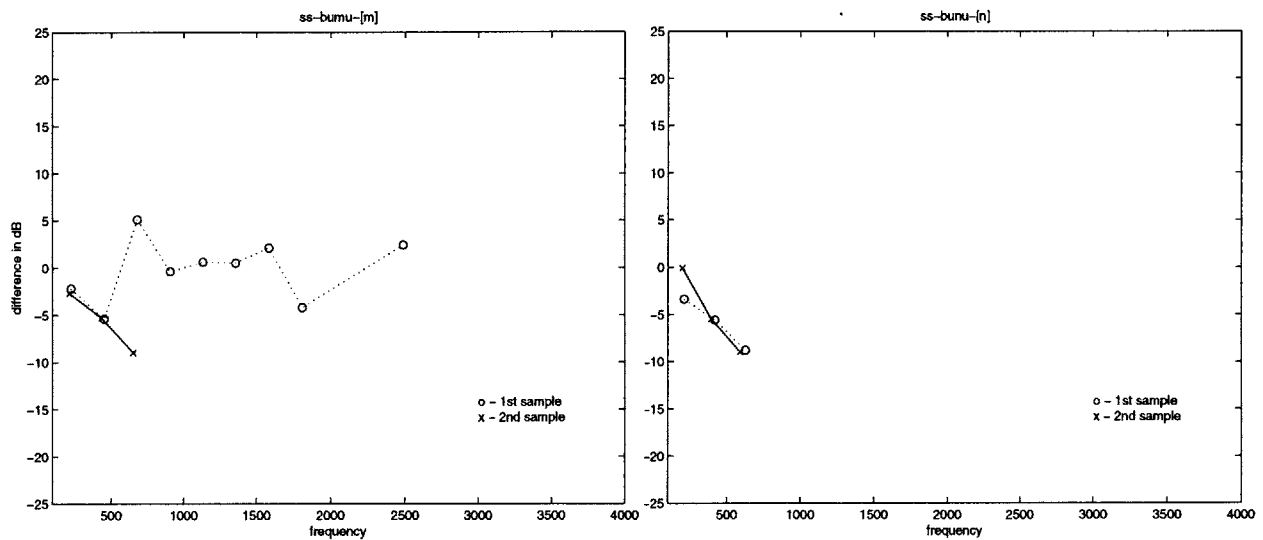
**Figure 5.7:** am [m] and [n] difference plots for bumu and bunu

The difference plots for am's [m] and [n] extracted from bumu and bunu are displayed in Figure 5.7. The only similarity that the two plots share is in the trajectory and values of the first three harmonic differences. Both plots have a first harmonic difference which is about +1 dB on both samples. The second harmonic difference drops to -3 or -4 dB and the third value decreases about one more decibel. However, after the third harmonic the [m] and [n] plots have distinctly different patterns, which is supported by the fact that for both plots the two samples are similar. Both have positive difference peaks but in the [m] plot it takes place around 1400 Hz while in the [n] plot it occurs around 1600 Hz. Moreover, these points may be anomalous due to the fact that in both the first sample does not display a similar pattern to the second.



**Figure 5.8:** ks [m] and [n] difference plots for bumu and bunu

The [m] and [n] difference plots in of ks have a little more similarity, at least on the first four points of the first sample of each plot (Figure 5.8). In both cases there is a decreasing in the value of the differences through the first three harmonic difference and then an increase of about 7 dB from the third to the fourth. The remaining portion of the plots look very dissimilar. The second sample of the [m] plot looks like the first sample. However, the second sample of the [n] plot does not look like its first. In fact it looks somewhat like the [n] plots of am. There was limited data in these plots again due to noise.



**Figure 5.9:** ss [m] and [n] difference plots for bumu and bunu

The difference plots for the [m] and [n] of ss are given in Figure 5.9. In all of the samples, except the first from the [m] plot, there are only three harmonic differences due to noise. The first three harmonic differences from the first sample of the [m] plot do not have the same pattern as the points in the second sample. The first two harmonic differences track well, but the third harmonic differences are very different. In the [n] plot the two samples look fairly similar except that the first point of the second sample is at 0 dB whereas the first point of the other sample is around -4 dB. However, both of these, but especially the first sample, look somewhat like the first sample of the [m] plot, so there seems to be some consistency there.

Overall the [m] plots in the context of bumu share some similarities. In both am's and ks' plots there is the similarity in pattern of the first four points. This same pattern can be seen in the [m] plot of ss. However, it is not exactly the same because of the fact her plot contains less harmonics. Nevertheless, her first two harmonic differences, in sample one, follow the pattern of the first

three harmonic differences of the two male speakers which have local minima at about 500 Hz. The third harmonic difference has a similar value, about 5 dB, and frequency location to the peak at the fourth harmonic difference in am's plot. In ss's plot the peak is located at about 600 Hz, where it is in ks's plot, and its appears to be at about 550 Hz in the plot of am. The [n] plots between speakers share almost no characteristics. The second sample of ks [n] somewhat resembles the first 5 points of those in the am [n] plot. The [n] plot of ss does not really resemble the [n] plots of either of the other speakers. Although the first two points form a segment down and to the right as do the first three points of the first sample in ks's [n] plot, the slopes are different. In addition, the third point of the ks plot, which would roughly correspond to the second point of the ss plot going by frequency, is a minimum but in the ss plot the difference continues to decrease on the next harmonic difference. In summation, there does seem to be some correlation between the nasals in the context, although the two samples of one plot do not always agree. However, at least one will match one sample of the same phone of another speaker. Thus, there may be a basic pattern for [m] in [n] within the context [u] in the lower frequencies between 100 to 750 Hz.

#### **5.1.4 Individual speaker results across all [m] and [n]**

The [m] difference plots of am showed almost no consistency at all. The [m] plot from bimi looked nothing like those of bama and bumu. The difference plots from bama and bumu have a vague resemblance in their first three points (except for the first sample of bama) and possibly an argument can be made for the fourth but the similarity ends there. Also, it appears that the first and second harmonic differences have a similar relationship in every [m] plot. Although there is not much consistency in the [m] plots there is a lot of consistency in the plots of [n] in the 100 to 600 Hz frequency range. All of the [n] plots share the same pattern for the first five harmonic dif-

ferences. The second harmonic difference is always about 5 dB below the first. The third and fourth harmonic difference stay within 1 dB of the second. The fifth harmonic difference rises 2 to 4 dB above the previous three differences, but never reaches the value of the first harmonic difference. The only thing which is consistent throughout both types of nasal plots is that all of the first harmonic differences are positive with the exception of the [m] plot from bama.

In the [m] plots of ks the constant was also the pattern of the first three or four harmonic differences. The first harmonic difference was always about 4 to 5 dB higher than the second. The change in the second to third was always only about 1 to 2 dB and the fourth always increased about 4 to 5 dB over the third except in bimi. The [n] plots shared similar characteristics as well. In fact, the shape and values of the first three harmonic difference was consistently similar throughout both type of nasal plots. The first harmonic differences in ks's plots were always negative except in the [m] plot from bimi. The traits of the first three harmonic differences are common to almost all of the ks nasal difference plots. There appears to be no special correlation between the vowel context and the resulting difference plots.

The [m] difference plots of ss have no consistency. The two samples in each plot varied and no two samples from any two plots are congruent in any way. The only exceptions are the second sample of the bama and bimi [m] plots but, the slope of the segment formed by the first and second points and the slope formed by the second and third points are distinctly different in each plot. However, it appears as if the angle between the two segments in each plot are similar which causes the plots to look alike. Each of the three [n] plots look similar in their first three harmonic differences, and in bana and bini, the fourth differences are similar was well. This may be the

case for bunu but the noise in the signal prevented extraction of data above the third harmonic. As was the case for ks, all of ss's first harmonic differences are negative with the exception of one.

### **5.1.5 Interspeaker [m] and [n] comparison**

Overall only ks had any type of consistency in his [m] difference plots. The other two speakers had very little consistency and because of this there are no signs of any type of correlation between the [m]'s of each speaker. Nor does there appear to be any correlation between the [m] difference plots in any vowel context, besides the two which were mentioned in the previous sections (the [m] plots for bama and bumu of am and ks). Although there was a lot of consistency between the [n]'s of each speaker, there was not much correlation between the [n] plots between speakers. In other words, am's [n] difference plots looked alike in pattern regardless of context as did those of ks, but when comparing the plots of ks and am there is not much parity. However, ks and ss did share a little similarity in the almost linearly decreasing of harmonic difference in through 500 Hz in many of their [n] difference plots. But above 500 Hz there were seldom signs of any similarity.

## 5.2 Analysis

On a few occasions the nasal difference plots between speakers had a similar pattern in the 100 Hz to 750 Hz frequency range. On even fewer occasions the difference plots of [m] and [n] from within the same vowel context of one speaker showed the symmetry which was predicted.

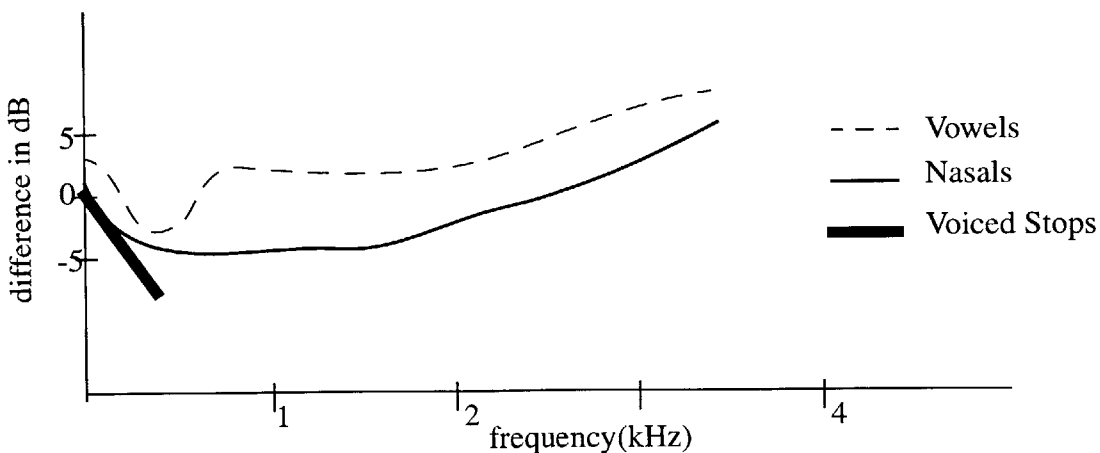
There is no solid evidence of consistency for nasals within the same context because not all speaker's plots exhibited consistency. However, there is evidence of consistency for the differences in the plots being solely dependent upon the type of nasal, [m] or [n]. And this phenomenon appears to be speaker dependent. Speaker dependency is logical because of the different anatomies of speakers in the face and neck region. The position of the nose openings with respect to the microphone will be different for each speaker because the face of each speaker is shaped differently. For some speakers the nose may protrude further out than others, the nostrils will be of different size, etc. Such variations in anatomy may be the reason for the differences between ks's and am's difference plots. In all of am's difference plots the minima were not as low in value as those for ks and the first harmonic different was almost always above 0 dB. It is as if the plots from am were shifted up the "difference in dB" axis with respect to ks's plots. The direction of sound propagation with respect to the close talking microphone may have been closer to the front of the microphone when am was speaking. This would mean that am's nasal speech signal were subject to less attenuation than ks's, and thus the difference between the spectra of the close-talking microphone subtracting the omnidirectional microphone would be less negative. Another aspect of the sound which may affect the manner in which it is transduced by the microphones is the fact that in the nearfield the sound waves will not be planar. They will be even less planar for

the close-talking microphone which is closer to the source. This affect can not be normalized and the difference it causes will be present in the spectra.



# 6 Conclusion and Recommendations

The difference plots of the vowels and nasal consonants exhibited certain patterns for each speaker. And, in general, the vowels and nasal consonants produced results that were somewhat expected. The vowels did not show large differences between the spectra of the two different microphones for frequencies below about 750 Hz. More specifically, the differences remained within plus or minus 5 dB over this frequency range. The nasal consonant difference plots exhibited mostly negative differences, which means the amplitude of the close-talking microphone was less than that of the omnidirectional, at least in the lower frequencies, which was also expected. Due to noise, there generally was not much data in the difference plots of the voiced stop consonants. However, there were a couple of trends exhibited in the plots as pointed out in section 4.2. There was always a negative decrease in the difference following the first harmonic difference, though the degree of this change in the difference was dependent upon the speaker. A sketch of the general trends for each phone type is given in Figure 6.1. The sketch shows a smooth curve of the trend in the differences on average for each phone type.



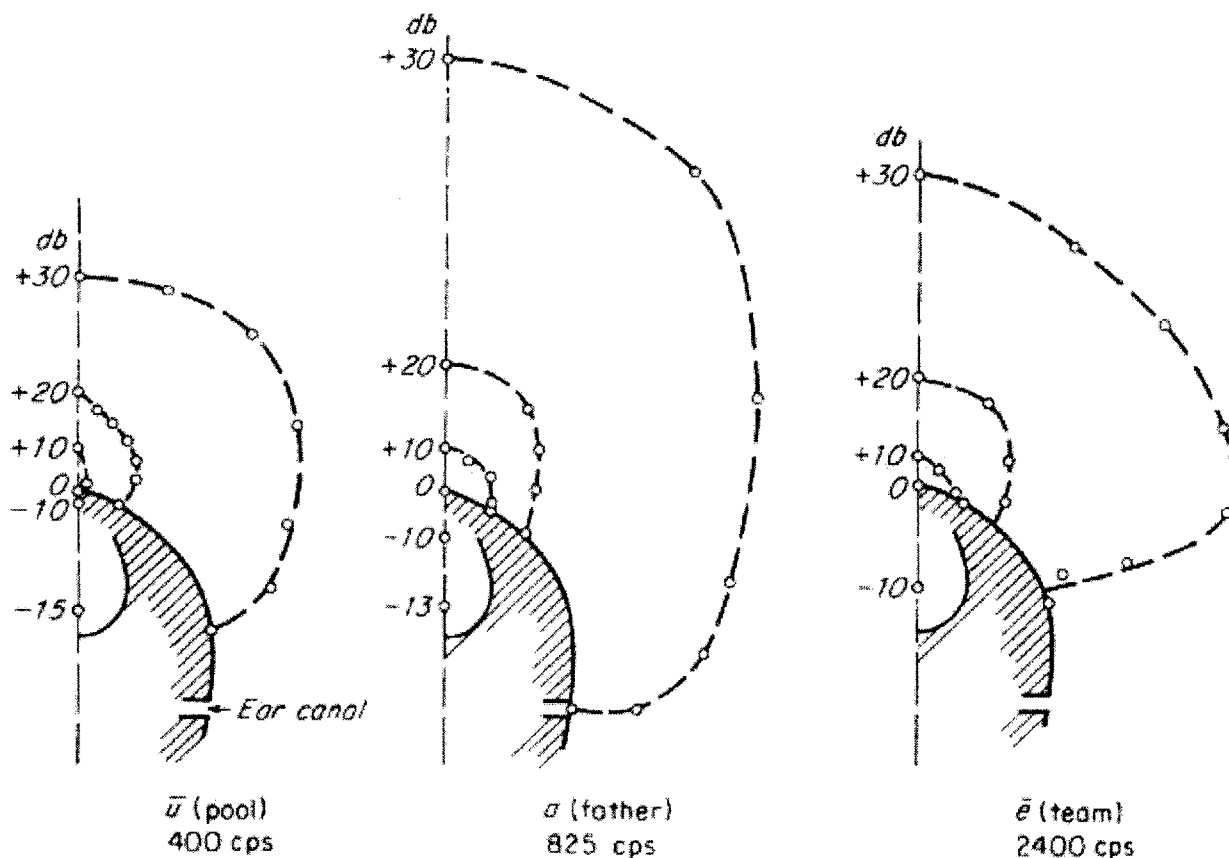
**Figure 6.1:** Sketch of general trends for each phone type

A result that was common to both the vowels and the nasal consonants was the local minima that appeared in all of the difference plots in the 100 Hz to 500 Hz frequency range. They were of varying “bandwidth”, but nevertheless each began with the first harmonic difference and the difference would go in the negative direction until about the third harmonic difference at which point it would the direction of the change in difference would be positive and continue until the fourth or fifth harmonic. This is mostly true for the two male speakers because they can have four to five harmonics in this frequency range. However, the female speaker, ss, also had these local minima in her difference plots. These minima may have been caused by a difference in the bandwidths of the first formants of the spectra produced by each microphone. If the peaks of two formants are matched and they have equal bandwidth, the difference between each corresponding point on both will be zero, or very close to zero. However, if the bandwidths are not equal and the peaks are matched the points of the formant with the narrower bandwidth will be of lesser value than the points on the other with a decreasing difference as the points get closer to the peak. This loss would have to be the result of a loss in amplitude due to some near field effect, or perhaps, and less likely, a conductive loss. The energy loss could not have been a result of the six decibel per doubling of distance loss that occurs in the far field.

Initially it was postulated, in section 3.2, that the first harmonic differences may have been more positive than the following two or three harmonic differences due to differences in the channel responses. But, if this were the case, it would be expected that it would have the same relation to the other harmonics the majority of the time, meaning it would always be about 5 dB more positive than the second harmonic difference instead of varying between 3 to 10 dB more positive of each speaker. In addition the contour of the low frequency minima would remain about the same

for each speaker across their vowels and nasal consonants. However this is not the case. Furthermore, the difference plots of ss also exhibited minima, and her first harmonic difference was always at a frequency at least 50 Hz above those of the two male speakers. So, it seems that the first harmonic differences' dominance over the next few differences is probably the cause of an acoustic near field effect like the one discussed in the previous paragraph.

A problem with both models for vowels and consonants is that they were derived for sound in the far field. Research has been conducted that shows the nearfield effect for vowel sounds about the head. An example of this is shown in Figure 6.2 [5] which displays the results of an experiment performed by Georg von Békésy that shows the magnitude response of three different vowels about the head is different.



**Figure 6.2:** Near field response about a coronal cross-section of the head for three vowels [5]

Now that the differences between the spectra of close-talking and omnidirectional microphones are understood a little better, maybe measurements which are not sensitive to microphone type and placement can be obtained and used in speech recognition. Therefore, if further study comparing the spectra of different microphones is to be performed other vowels and consonants must be examined. In addition, the following recommendations might also be helpful for more control in the research:

- Better noise elimination from equipment.
- Use the same preamplifier for more of an equal channel response, or test each microphone in both channels to find any differences in the channel response.

- Measure the nearfield response about each speakers head in an anechoic chamber, also determine where the nearfield ends.

- Place the close-talking microphone in several position around the mouth to determine how much the placement matters for particular sounds.

- Take several pictures of each speaker's face/head to understand their anatomy better and to better account for the variations in the differences of the spectra of each speaker.

## **Acknowledgments**

First and foremost I would like to thank Professor Ken Stevens for the idea that became this thesis project. Also, and most importantly, I would like to thank him for his patience and guidance throughout the year in helping to see me through this thesis project. In addition I would like to thank him for being a speaker for the research.

I would also like to thank Janet Slifka, Harold Cheyne, Aaron Maldonado, and Stefanie Shattuck-Hufnagel for all agreeing to be speakers in the research.

My sincere appreciation goes out to Arlene Wint for all of her help in the administrative issues that had to be dealt with throughout the year.

## References

- [1] John M. Eargle, Handbook of Recording Engineering, New York, New York: Van Nostrand Reinhold, 1992.
  
- [2] Sennheiser electronic GmbH & Co., “HMD 414-6 Technical Data” pamphlet, Germany, 1984.
  
- [3] Kenneth N. Stevens, Acoustic Phonetics, Cambridge, MA.: The MIT Press, 1998.
  
- [4] Shure Brothers Inc., “M68 Microphone Mixer Professional Products Data Sheet”, Evanston IL., 1979.
  
- [5] Georg von Békésy, Experiments In Hearing, New York, New York: McGraw-Hill Book Company Inc., 1960.