Measuring the Neck Transfer Function of Laryngectomy Patients

by

Geoffrey S. Meltzner

B.S, Engineering (1995)

Harvey Mudd College

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

at the

Massachusetts Institute of Technology

February 1998
[February 1999]

Signature of Author ....
Department of Electrical Engineering and Computer Science
January 15, 1998

Certified by ....
Kenneth N. Stevens
Professor of Electrical Engineering
Thesis Supervisor

Certified by ....
Robert E. Hillman
Associate Professor of Otology and Laryngology, HMS
Thesis Supervisor

Accepted by ....
Professor Arthur C. Smith
Chairman, Department Committee on Graduate Students

Measuring the Neck Transfer Function of Laryngectomy Patients

By

Geoffrey S. Meltzner

Submitted to the Department of Electrical Engineering
and Computer Science on Jan. 15, 1999 in partial fulfillment of the
requirements for the Degree of Master of Science in
Electrical Engineering

Abstract

Measurements of the Neck Transfer Function (NTF), defined as the ratio of the estimated volume velocity that excites the vocal tract to the acceleration measured at the neck wall when the neck is driven by a mini-shaker, were made at three different positions on the necks of ten laryngectomized subjects (five males and five females) and four normal laryngeal speakers (two males and two females). A mini-shaker driven by broadband noise, provided excitation to the necks of subjects as they configured their vocal tracts to mimic the production of the vowels /a/, /ae/, and /I/. The sound pressure at the lips was measured with a microphone and an impedance head mounted on the shaker measured both the force and acceleration.

Overall, the results indicated that the neck wall passes low frequency sounds better than high frequency sounds. A comparison of the NTFs of four subject sub-groups revealed that the NTFs of female laryngeal speakers differed from the NTFs of the other groups (male laryngeal speakers, laryngectomized males and laryngectomized females). In addition, there was a notable amount of inter-subject variability even within a subject group. The results also indicate that the NTF is relatively independent of the amount of pressure with which one applies the shaker to the neck. These results should be useful for aiding in the design of an improved neck-type electrolarynx.

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

Thesis Supervisor: Robert E. Hillman
Title: Associate Professor of Otology and Laryngology, Harvard Medical School

Acknowledgments

This work would never have been completed with out the help of a great many people who I would like to thank. These people have assisted in practically every aspect of this research. First, I'd like to thank Bob Hillman for taking me on as a student and giving me the opportunity to work on the artificial larynx project, of which this research is a part. He has also provided a great deal of support and guidance that have helped me throughout this project.

I'd also like to thank Ken Stevens for kindly agreeing to supervise me on this thesis and sharing his expertise with me so that I could improve the quality of my research.

Jim Kobler has also been of great assistance in helping me construct some of the experimental equipment, helping me refine the experimental protocol and for assisting in performing the experiments.

I'd also like to thank Ken Houston at Draper Labs for providing his expertise and experimental equipment, and Ehab Goldstein for helping me run several of the experiments.

Glenn Bunting and Janice Eng at the Voicelab also deserve my thanks for recruiting the laryngectomy patients that I needed to study.

Harold Cheyne deserves a hearty thanks for letting me bounce ideas off of him and for acting as a short-notice subject.

Finally, I am greatly indebted to John Rosowski at the Eaton Peabody Lab. John has acted as an unofficial third advisor, always willing to answer my questions, no matter how often I pestered him. He provided me with experimental equipment I needed, helped me numerous times with various tasks. Without John's help, this project would never have gotten off the ground.

Table of Contents

I. Introduction

Each year, thousands of people lose the ability to speak because they are laryngectomized or suffer laryngeal trauma. These people must rely on alternative methods for producing speech, often referred to as alaryngeal speech. Some alaryngeal speakers are able to learn esophageal speech, while others use tracheoesophageal (T-E) speech. However, because esophageal speech is difficult to learn and because the anatomy of some laryngectomy patients prevent them from using T-E speech, a large percentage of alaryngeal speakers rely on an artificial larynx, also known as the electrolarynx. Moreover, many esophageal and T-E speakers also use electrolarynges as an intermediate step until they are able to master other forms of alaryngeal speech (Hillman *et al.* 1998).

The electrolarynx is available in two forms: the intra-oral (or transoral) form and the transcutaneous (or transcervical or neck-type) form. Both are hand-held devices that vibrate with a periodic signal. The intra-oral version includes a small tube which is placed in the mouth to excite the vocal tract, whereas the transcutaneous version is held against the neck to drive the vocal tract externally. Because this thesis focuses on the transcutaneous electrolarynx, any use of the term "electrolarynx," unless otherwise specified, will refer to this version of the device.

Electrolaryngeal (EL) speech generally provides a serviceable means of communication, but it does have several shortcomings. EL speech has an artificial quality that often attracts unwanted attention to the user and in certain situations, such as in a crowd, EL speech is often not loud enough to be heard. EL speech also has a reduced intelligibility, especially when trying to discriminate between voiced and unvoiced consonants (Weiss and Basili 1985, Weiss *et al.* 1979).

Despite being developed almost forty years ago (Barney *et al.* 1959), there have been few efforts to improve the electrolarynx. Qi and Weinberg (1991) attempted to improve the quality of EL speech by enhancing its low frequency content and claimed that the enhanced speech was judged to be more pleasing than the speech without enhancement. In an effort to improve the intelligibility of EL speech, Chari *et al.* (1997) used an adaptive filtering algorithm to remove the contribution of the directly radiated sound (produced by the EL) to the perceived speech signal. This processing resulted in increased intelligibility for stop consonants but degraded nasal consonant intelligibility.

Norton and Bernstein (1993) tried to improve EL speech by changing the driving signal of the vibration source. They measured the frequency response function (FRF) of the neck, which they defined as the ratio of the pressure signal measured at the lips (after the formants have been removed) to the input signal of a shaker which vibrated against the neck. They used the neck FRF to modify the output of an electrolarynx such that the spectrum of the sound that excited the vocal tract resembled that of a natural human glottal source. Listeners informally judged the speech produced using their modified electrolarynx to sound more natural than speech produced using commercially available electrolarynges.

While the results of the Norton and Bernstein study are promising, there were issues that were not addressed. Norton and Bernstein measured the frequency response of the neck of a single normal subject, but the anatomy of the necks of laryngectomy patients differs significantly from normal necks. Necks of laryngectomees lack the larynx and its surrounding tissue, and additionally, laryngectomy patients are often subjected to radiation treatment, which can change the properties of the neck tissue. These differences could also produce an FRF that is very different from the one measured in a normal neck.

Norton and Bernstein's work also does not address the variations in the FRF that may occur from patient to patient. The surgery performed on one patient tends to differ from that performed on another due to differences in the size and location of the tumors and the involvement of other neck structures. These differences in surgical techniques could potentially affect the transmission properties of the neck wall.

Finally, Norton and Bernstein only made a measurement on a single place on the neck. However, because of the asymmetric nature of a laryngectomy surgery, especially if a radical neck dissection follows the laryngectomy, the characteristics of the tissue at different locations on the neck may differ. Furthermore, EL users often claim that there is only one spot on their neck which produces speech loud enough to understand. Therefore, it seems likely that the neck transfer function will vary with neck position, especially for laryngectomy patients.

Thus, the goal of this thesis was to extend the Norton and Bernstein study by measuring the neck transfer functions of both laryngectomized and non-laryngectomized subjects, at different locations on the neck. With this information, one could theoretically design an improved electrolarynx driving signal which would excite the vocal tract in a manner similar to that of a natural glottal sound source. Furthermore, if a large degree of variation between the neck transfer functions of different subjects is discovered, then the neck transfer function can be used to tailor an electrolarynx to its user.

II. Procedure

The following experimental procedures were used to collect the necessary data to determine the neck transfer function (NTF). The NTF is defined as the ratio of the estimated volume velocity that excites the vocal tract to the acceleration measured at the neck wall when the neck is driven by a mini-shaker. These procedures are based on the ones used by Norton and Bernstein (1993) and Fujimura and Lindqvist (1971) but have been modified to improve the accuracy of the measurements.

A. Subjects

The experimental subject group consisted of 10 laryngectomized subjects (5 males and 5 females) all of whom had been laryngectomized at least 1 year prior to the date of the experiment, and were given post-surgical radiation treatment. The subjects ranged between 50 and 76 years in age (mean of 65.6 ±7.8 years). All of the laryngectomized subjects were patients at the Voice and Speech Laboratory at the Massachusetts Eye and Ear Infirmary. All but two laryngectomized subjects were part-time or full-time electrolarynx users

In addition to the laryngectomized subjects, 4 normal subjects (i.e. non-laryngectomized, or laryngeal speakers), 2 male and 2 female, were used. The ages of the normal laryngeal speakers ranged from 22 to 31 (mean 25.8±3.9 years). Although none of the normal subjects was an everyday electrolarynx user, all of them were familiar with how to use the device.

B. Experimental Protocol

All of the experiments were performed in the sound treated experimental chamber in the Voice and Speech Lab at the Massachusetts Eye and Ear Infirmary. Two separate experiments were performed. The first, main experiment was performed on the 10 laryngectomized and laryngeal subjects described in the preceding section. A second experiment was performed on a single laryngeal male subject in order to estimate the near field lip radiation characteristic. Section II-D explains the need for this second experiment.

1. Main Experiment Protocol

1. Each subject was seated in a clinical exam chair and the chair was adjusted to make the subject comfortable.
2. A piece of headgear with a small, directional microphone (Sennheiser) mounted to it was placed on each subject's head. This headgear is used in a similar fashion for routine acoustic measurements made by clinicians in the Voicelab. The microphone location was adjusted such that it was placed 1 cm away from the subject's lip and was oriented so that it faced the lips. The microphone was locked into place to ensure that it remained the same distance from the subject's lips throughout the entire experiment. The microphone was calibrated using a Bruel and Kjaer ¼" microphone which was assumed to have a flat response for $60 < f < 4000$ Hz.
3. A small Bruel and Kjaer mini-shaker (model 3081), mounted with a 2" long cylindrical shaft and an impedance head (a transducer that simultaneously measures force and acceleration – PCB Model 288D01/788D01) and a small 2.5 cm. diameter cap was placed on the subject's neck. The size of the cap is the same as that found on a Servox electrolarynx.
4. The shaker was driven with broadband noise (BW=12500 Hz).
5. While the shaker was being the driven, the subject was asked to configure his. her vocal tract as if he/she were saying the vowels /a/ (as in cot), /ae/ (as in cat), and /I/ (as in kit). Each vowel was done as a separate trial. In an additional trial, the subject was asked to keep

his/her mouth closed while the shaker was being driven. All of the non-laryngectomized subjects were instructed to maintain the glottis in a closed position by holding his/her breath. The vocal tract with a closed glottis simulates the laryngectomized vocal tract in which the vocal tract is decoupled from the subglottal system and is rigidly terminated.

6. As the subject performed the experimental task, a data acquisition system (Axon Instruments' Digidata acquisition board with accompanying Axoscope software) simultaneously recorded the pressure signal from the microphone, the force and acceleration signals produced by the impedance head, as well as the voltage of the shaker driving signal. All of the signals were sampled at 25 kHz after being low pass filtered at 6 kHz by 4-pole Bessel filters (Axon Instruments Cyberamp).
7. In an additional task, the subject was asked to change his/her vocal tract configuration as if he/she were saying the vowels, /a/, /ae/, and /I/ in continuous succession, while the shaker was being driven.
8. Steps 5 – 7 were performed at three locations on the neck. The first position was always the location where the subject placed his/her electrolarynx during everyday use. For those subjects who did not use an electrolarynx, position 1 was arbitrarily chosen to be a spot on the upper right neck. Position 2 was chosen to be a position 2 cm. below position 1, and position 3 was the mirror image of position 1 on the opposite side of the neck. For some laryngectomized subjects, measurements at all three positions were not obtained because at these locations the signal transmission was too weak to obtain any meaningful data.
9. Finally, the shaker was placed back in the first tested position and the subjects were asked to once again configure their vocal tracts to mimic the production of the vowel /a/ while the shaker was driven. This task was repeated three times; However with each successive trial,

the amount of DC force was increased. The force transducer on the impedance head was configured to simultaneously measure both AC and DC Force (on two separate channels) and in this case, the DC Force channel measured the amount of force each subject used to hold the shaker in place. This task was used to collect data to determine if changes in DC Force affect the NTF.

10. The entire experimental session was video taped, thereby allowing the monitoring of the position of the shaker on the neck to ensure that it was not moved during each experimental trial.

A schematic of the experimental set-up is shown in Figure 2.1.

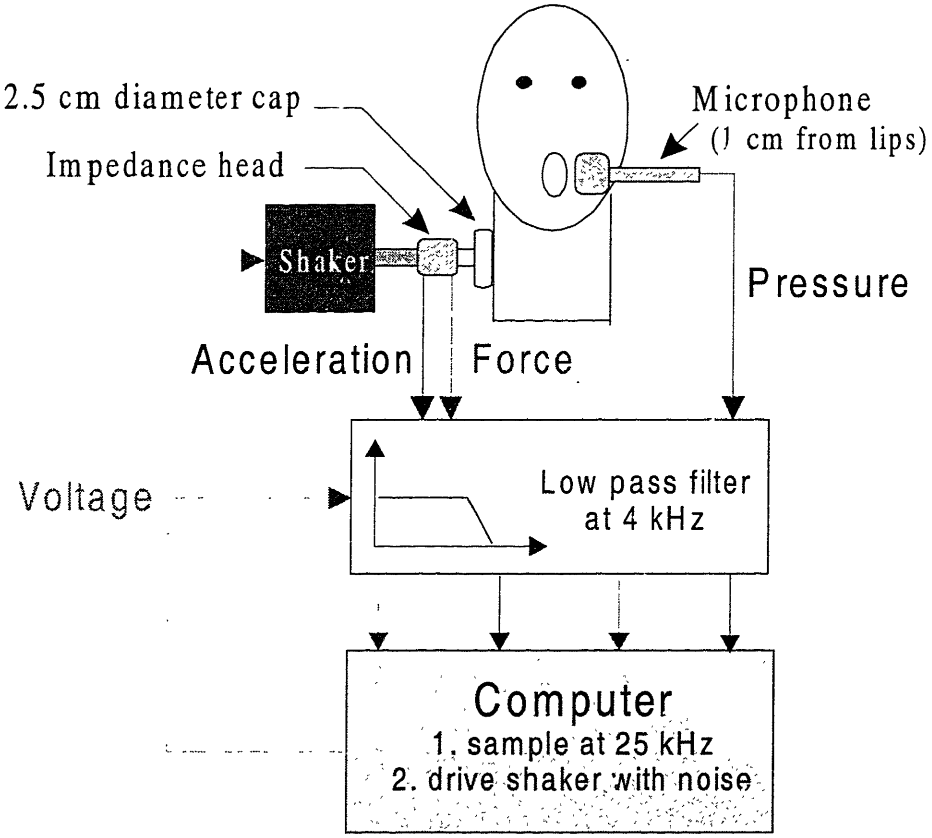


Figure 2.1. Schematic of the experimental set-up for collecting data to measure the neck transfer function

2. Secondary Experimental Protocol

As in the main experimental protocol, the subject (a laryngeal male, age 24) was seated in the adjustable chair in the sound booth and fitted with the headgear upon which the microphone was mounted. The headgear was adjusted such that the microphone was placed 1 cm away from the subject's lips. An additional microphone was suspended from the ceiling at a distance of 45 cm from the subject's lips. The subject was asked to sustain the vowels /a/, /ae/, and /I/, for a duration of 5 seconds. Finally, the microphone mounted on the headgear was then moved away from the lips and the subject was again asked to sustain the same three vowels.

C. Technical Issues

1. Directly Radiated Sound from the Shaker

As is discussed in Section II-D, the microphone at the lips is not only measuring sound that is filtered by the neck and the vocal tract, but also the sound that is transmitted directly from the shaker to the microphone via an airborne path. At certain frequencies, this directly radiated sound is as at least as intense as that which exits the subject's lips (see Section IV-D), thus confounding the estimation of the neck transfer function. Therefore it became necessary to find a method of attenuating the direct radiation. The shaker was placed in a large plastic cylinder that was packed with sound attenuating foam. The cylinder was fitted with a cap that contained a ½" hole in the middle so that the cylindrical shaft on which the impedance head is mounted could move freely. Figure 2.2 shows a schematic of the shaker encased within its silencer. The silencer proved to be effective especially for frequencies between 600 and 4000 Hz. Figure 2.3 presents the spectra of the direct radiation before and after the shaker is placed in the silencer.

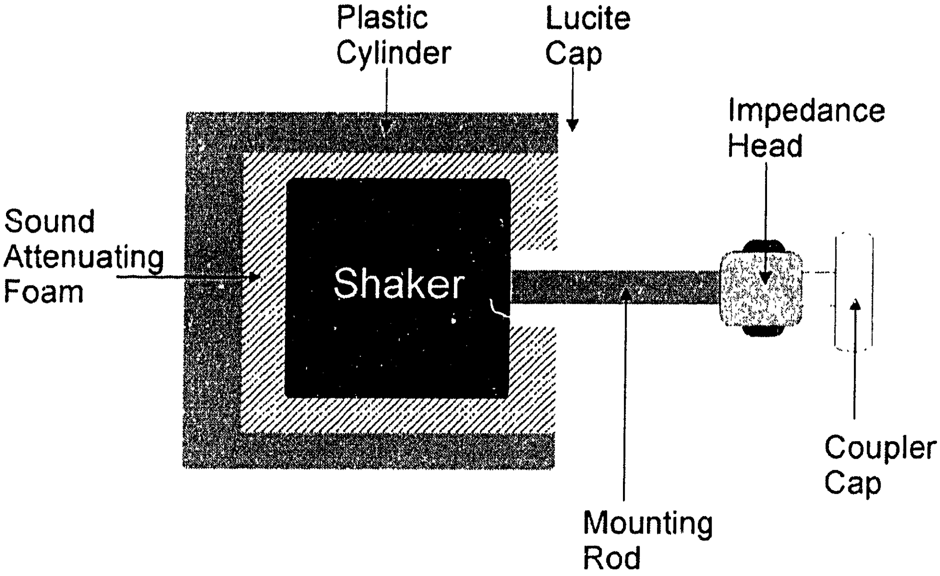


Figure 2.2. Schematic of the shaker enclosed inside the cylinder. Special care was taken to make sure that the shaft upon which the impedance head is mounted was able to vibrate freely so as not to affect the input signal.

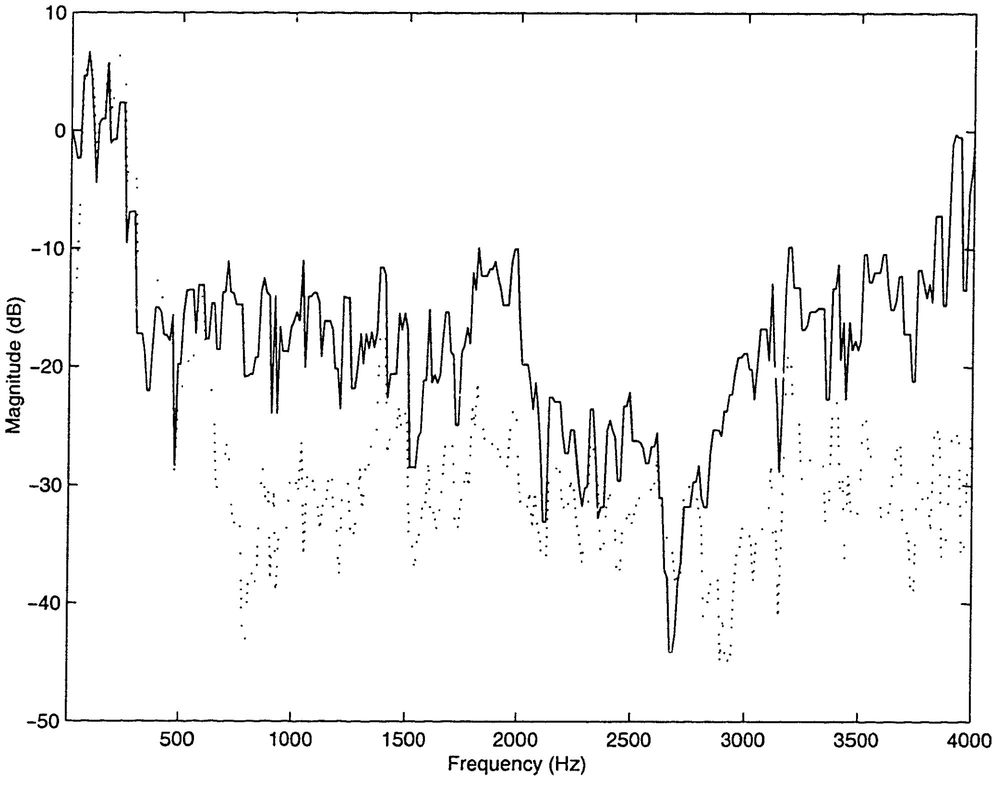


Figure 2.3 Spectra of the direct radiation of the shaker both in (dotted line) and out (solid line) of the cylindrical silencer. These measurements were made by recreating the closed mouth measurements described in the previous section. Although the silencer is not as effective at frequencies below 600 Hz, it does perform very well at high frequencies, sometimes attenuating the sound but as much as 20 dB.

2. Mounting the shaker

Although the shaker itself is not that large (about 2" in diameter), it is quite heavy. Holding it against the neck for an extended period of time can fatigue the subject and cause him/her to change the position of the shaker on the neck, or move the shaker during the experimental task. The shaker becomes more unwieldy when placed in the silencer since the cylinder became about 5" in diameter. To make the shaker easier to handle, the silencer was fitted with a counterweight and suspended by a rope from the ceiling of the experimental chamber. With the shaker mounted in this fashion, the majority of the weight of the shaker is supported by the ceiling and holding the shaker in place requires minimal effort.

D. Data Analysis

1. Estimating the Neck Transfer Function - Theory

The transfer function that was determined is the ratio of the signal that excites the vocal tract (i.e. volume velocity) to the input acceleration. However, while the acceleration is directly measured, some signal processing must be done to obtain an estimate of the volume velocity from the pressure signal measured at the mouth. To illustrate the signal processing algorithm that was used, it will be helpful to use the block diagram displayed in Figure 2.4.

In this model it is assumed that each block represents a linear, time-invariant system. If this block diagram is used to model the system in question, it can be treated as if there are two paths. In the first path, the driving voltage signal (broadband noise with an 12.5 kHz bandwidth) is transformed into an acceleration signal by the shaker; this transformation is characterized by the primary shaker transfer function, $T_1(f)$. The acceleration is then filtered by the neck wall ($N(f)$), vocal tract ($VT(f)$), and the near field lip radiation characteristic ($R_n(f)$).

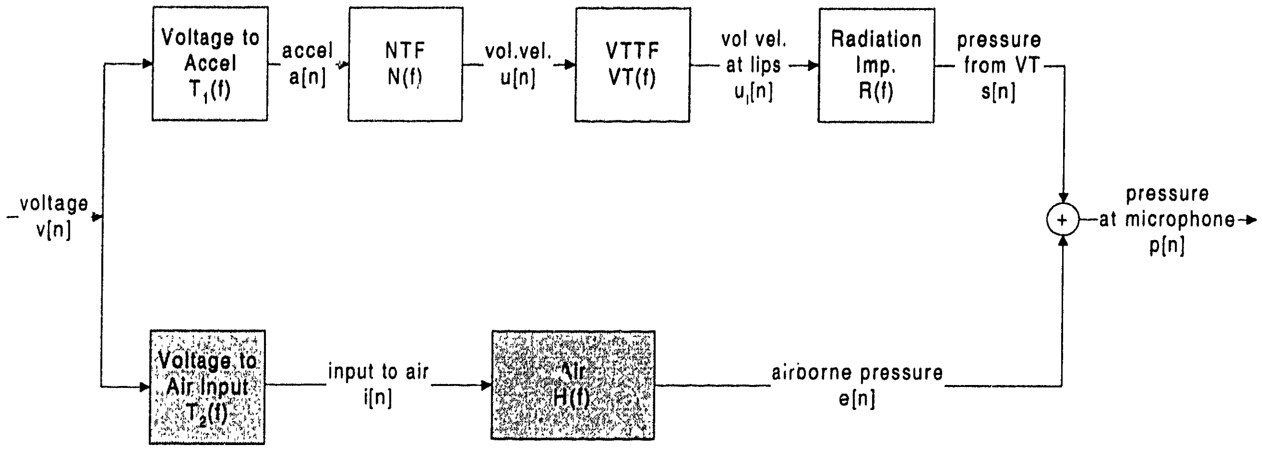


Figure 2.4. Block diagram of experimental setup. The microphone at the lips measures the sound that is the sum of the sound transmitted through the neck wall and vocal tract and the sound transmitted via a direct path through the air. It should be noted, however, that although the measured signals were continuous, to make use of the analysis algorithm, they needed be digitized, and therefore, the discrete signal notation $x[n]$ will be employed where $x[n]$ is the discrete counterpart to $x(t)$.

The output of this path is represented by $s[n]$ which is the pressure measured by the microphone if this path could be isolated from the rest of the system. In the second path, the voltage signal is transformed into a signal which excites the air by a secondary shaker transfer function, $T_2(f)$. Although the exact nature of this transfer function is unknown, its effects will be effectively removed from the signals that will be analyzed and therefore it is only important that it is a linear time-invariant system. The validity of this LTI assumption is discussed in the next section. The input to the air is then filtered by an unknown air transfer function, $H(f)$, (also assumed to be LTI) and the resulting pressure, $e[n]$, is measured by the microphone located at the lips when the mouth is closed. It should be noted that $H(f)$ also includes any filter characteristics due to the directionality of the microphone.. The four signals that were measured are $v[n]$ (the voltage driving the shaker), $a[n]$ (the acceleration signal from the accelerometer), $p[n]$ the pressure signal from the microphone when the lips are open, and $e[n]$, the pressure signal from the microphone when the lips are closed, where

$$p[n] = s[n] + e[n]. \qquad (1).$$

However, the transfer function to be estimated is $N(f)$ where

$$N(f) = \frac{S_{ua}(f)}{S_{aa}(f)}, \qquad (2)$$

and $S_{ua}(f)$ and $S_{aa}(f)$ are the cross-spectral density of $u[n]$ and $a[n]$ and the power spectral density of $a[n]$ respectively. (It should be noted that since the input to the system is a random process (broadband noise), that the transfer functions must be determined using cross and power spectral densities.) But, since $v[n]$, $u[n]$, $p[n]$, and $e[n]$ are the signals that were measured, $N(f)$ cannot be computed directly. Therefore an alternative method for computing $N(f)$ needed to be developed. The resulting algorithm is discussed below.

Step One – Removing the Airborne Path

Given the relationship described in equation (1), the following equation is also valid:

$$R_{pv}[n] = R_{sv}[n] + R_{ev}[n] \qquad (3)$$

where $R_{pv}[n]$, $R_{sv}[n]$, and $R_{ev}[n]$ are the cross-correlations of $v[n]$ with $p[n]$, $s[n]$ with $v[n]$, and $e[n]$ with $v[n]$ respectively. Furthermore, because each system block is assumed to be an LTI system,

$$R_{sv}[n] = R_{vv}[n] * (t_1[n] * vt[n] * n[n] * r[n]) \text{ and,} \qquad (4a)$$

$$R_{ev}[n] = R_{vv}[n] * (t_2[n] * h[n]) \qquad (4b)$$

where $R_{sv}[n]$ is the cross-correlation of $s[n]$ and $v[n]$, $R_{ev}[n]$ is the cross-correlation of $e[n]$ and $v[n]$, and $R_{vv}[n]$ is the autocorrelation of $v[n]$. Combining equations (3) and (4a,4b) results in

$$R_{pv}[n] = R_{vv}[n] * [t_1[n] * vt[n] * n[n] * r_n[n] + t_2[n] * h[n]]. \qquad (5)$$

Performing a Fourier transform on equation (5), dividing by $S_{vv}(f)$ (the power spectral density of $v[n]$), and rearranging terms leads to

$$\frac{S_{pv}(f)}{S_{vv}(f)} - T_2(f)\cdot H(f) = T_1(f)\cdot VT(f)\cdot N(f)\cdot R_n(f) \tag{6}$$

where $S_{pv}(f)$ is the cross-spectral density of $p(t)$ and $v(t)$. But, transforming equation (4) into the frequency domain produces

$$\frac{S_{ev}(f)}{S_{vv}(f)} = T_2(f)\cdot H(f) \tag{7}$$

and therefore,

$$\left(\frac{S_{pv}(f)}{S_{vv}(f)} - \frac{S_{ev}(f)}{S_{vv}(f)}\right)\cdot\frac{1}{T_1(f)\cdot VT(f)\cdot R_n(f)} = N(f). \tag{8}$$

Thus, if $S_{ev}(f)$, $T_1(f)$, $VT(f)$, and $R_n(f)$ can be computed, then equation (8) can be used to determine the Neck Transfer Function, $N(f)$.

To compute $S_{ev}(f)$, it is necessary to measure $e[n]$. If the lower pathway in Figure 2.4 could be isolated from the entire system, then this measurement is trivial. The simplest way to isolate this pathway is to make a recording while the subject's mouth is closed. In this task, no sound (ideally) is radiating from the lips and thus $e[n]$ can be estimated as $\hat{e}[n] = p[n]$. Therefore, $\hat{S}_{ev}(f)$ and $\hat{S}_{vv}(f)$ are computed from $\hat{e}[n]$ and $v[n]$ as estimates of $S_{ev}(f)$ and $S_{vv}(f)$ respectively. Thus, by using the data from the closed mouth measurements with equations (6) and (7), the corruption introduced to $s[n]$ by the direct airborne pathway is effectively removed. The remaining steps in the analysis algorithm involve computing $T_1(f)$, $VT(f)$, and $R_n(f)$ in order to isolate $N(f)$ from the rest of the system.

Step Two – Calculating the primary shaker transfer function.

After removing the effects of the direct pathway from the signals to be analyzed, the remaining transfer function, $K(f)$ is a cascade of four transfer functions, i.e.

$$K(f) = T_1(f) \cdot N(f) \cdot VT(f) \cdot R(f) \tag{9}$$

and hence $N(f)$ cannot be determined unless $T_1(f)$, $VT(f)$, and $R(f)$ can also be determined. Of the three confounding transfer functions, the simplest to calculate is the primary shaker transfer function $T_1(f)$. According to the block diagram in Figure 2.4, the input to $T_1(f)$ is $v[n]$ and its output is $a[n]$, and therefore $T_1(f)$ can be computed as

$$T_1(f) = \frac{S_{av}(f)}{S_{vv}(f)} \tag{10}$$

where $S_{av}(f)$ is the cross-spectral density of a[n] and $v[n]$. Because both $a[n]$ and $v[n]$ are measured during the experiments, $T_1(f)$ can be simply determined by measured signals. Figure 2.5 displays the magnitude and angle of this transfer function.

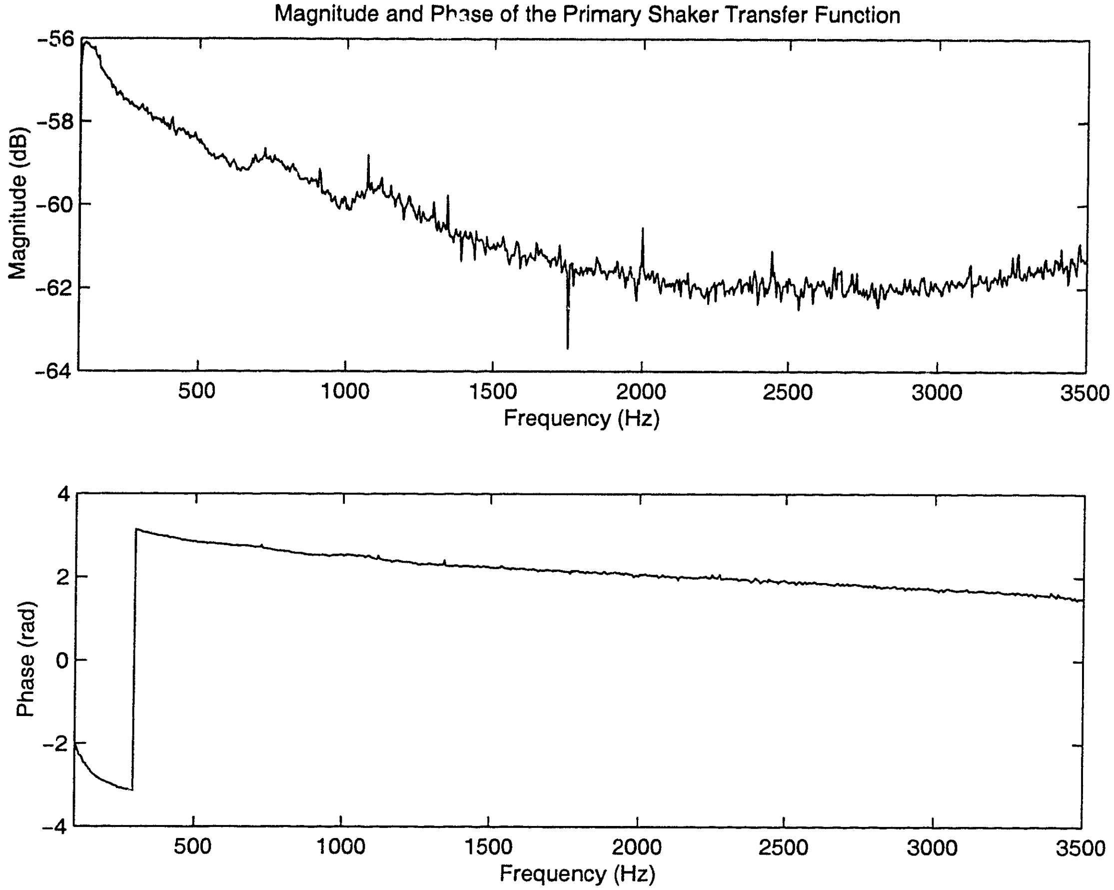


Figure 2.5. Magnitude and Phase of the Primary Shaker Transfer Function, $T_1(f)$ for the frequency range $100 < f < 3000$ Hz (the frequency range reported for the Neck Transfer Function). This transfer function was computed using data collected from one trial performed on a laryngectomized male subject.

Step Three: Estimating the near field lip radiation characteristic.

As shown in the block diagram of Figure 2.4, the output of the vocal tract (at the lips) is a volume velocity, $u_l[n]$. However, with the microphone located 1 cm from the lips, it produces a voltage that is proportional to pressure. Therefore in order to accurately estimate the neck transfer function, it is necessary to determine the relationship between the volume velocity at the lips, $u_l[n]$, and the pressure at the microphone, $s[n]$. This relationship is called the lip radiation characteristic, $R(f)$. If $U_l(f)$ is the spectrum of the volume velocity at the lips and $P_r(f)$ is the spectrum of the pressure at a distance r from the lips, then $R(f)$ is defined as

$$R(f) = \frac{P_r(f)}{U_l(f)}. \tag{11}$$

For distances, r, which are greater than a few centimeters, the open mouth can be considered a simple source radiating in all directions. In such cases, the radiation characteristic can be approximated as

$$R_f(f) = \frac{j2\pi f\rho}{4\pi r} \cdot e^{-j\frac{2\pi fr}{c}}, \tag{12}$$

where c is the speed of sound, and ρ, is the density of air. Equation (12) is an accurate approximation (within a few decibels) for frequencies up to 4000 Hz. (Stevens 1998).

In the NTF measurement experimental protocol, however, the microphone is placed 1 cm away from the lips and cannot be considered to be located at a large enough distance from the lips for equation (12) to be valid. Thus an alternative expression must be sought for the near field radiation characteristic. To this end, the second experiment described in the preceding section was performed. In this experiment, the pressure is measured simultaneously at a distance $r_1 = 1$ cm, $P_1(f)$, and at a distance $r_2 = 45$ cm, $P_2(f)$, from the lips. By using equation (11) with the measured value of $P_2(f)$, $U_l(f)$ can be computed as

$$U_l(f) = \frac{P_2(f)}{R_f(f)}. \tag{13}$$

Given that both $U_l(f)$ and $P_1(f)$ are now known quantities, substituting equation (13) into equation (11) produces an expression for the near field lip radiation characteristic, $R_n(f)$:

$$R_n(f) = \frac{P_1(f) \cdot R_f(f)}{P_2(f)} \tag{14}$$

and finally, if equation (12) is combined with equation (14), then $R_n(f)$ can be calculated as

$$R_n(f) = \frac{P_1(f)}{P_2(f)} \cdot \frac{j2\pi f\rho}{4\pi r} \cdot e^{-j\frac{2\pi fr}{c}} \tag{15}$$

Using the data collected when $P_1(f)$ and $P_2(f)$ were measured for a laryngeal subject saying the vowels /a/, /ae/, and /I/., $R_n(f)$ was computed for the shape of the mouth associated with each vowel. The magnitude of $R_n(f)$ for each vowel as well as a fifth order polynomial fit to each curve is shown in Figure 2.6. Because only the magnitude of the neck transfer function is reported, only the curve produced by the polynomial fit of the magnitude of $R_n(f)$ was used in the analysis.

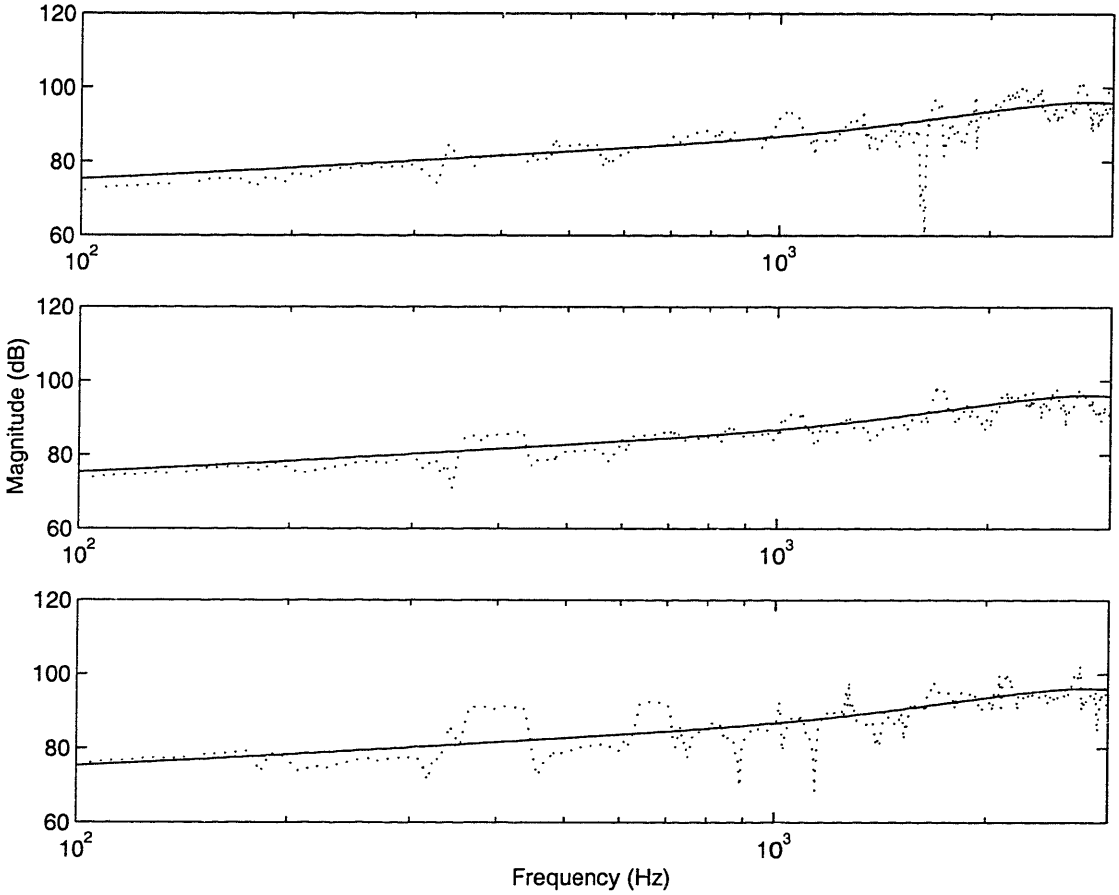
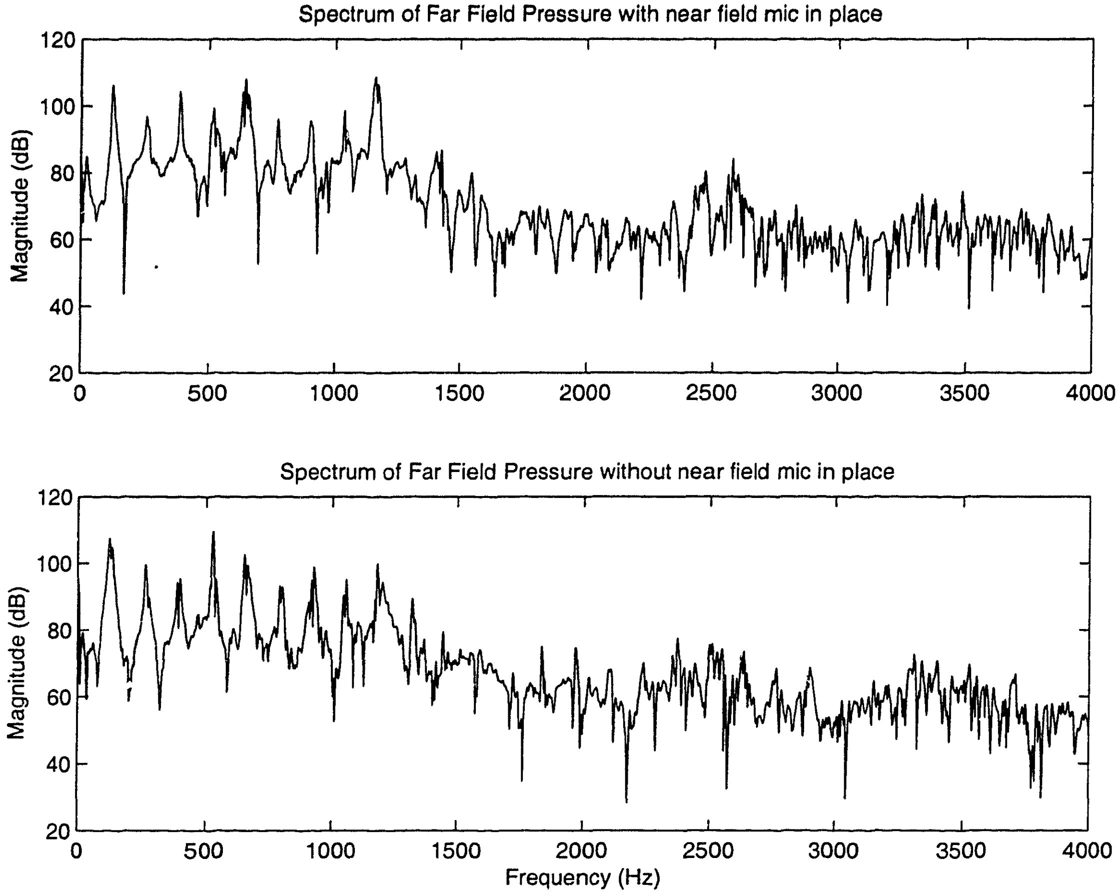


Figure 2.6. Computed magnitude (dashed line) and corresponding polynomial fit (sold line) for the lip radiation characteristic for the vowels /a/ (top), /ae/ (middle), and /I/ (bottom). Although there is some difference between the radiation characteristics for each vowel, the same polynomial could be used for all three vowels. These radiation characteristics were computed using c =331.6 m/s and ρ = 1.293 kg/m^3 (Kinsler, Frey, Coppens, and Sanders 1982).


Spectrum of Far Field Pressure with near field mic in place

Spectrum of Far Field Pressure without near field mic in place

Figure 2.7. Plots of the spectra of the far field pressures measured with (top) and without (bottom) the near field microphone in place. Both spectra look very similar and therefore it is unlikely that the near field microphone has a large effect on the measured far field pressure.

One possible problem associated with this experiment is that the presence of the near field microphone had an effect on the pressure measured in the far field. Although the near field microphone was very small, placing it so close to the source of the sound pressure measured at the far field could affect the measured spectrum in the far field. If the effect is significant, it could alter the computed near field radiation characteristic. To investigate the effect of the near field microphone, the subject was instructed to sustain the three vowels /a/, /ae/, and /I/ again, but this time the near field microphone was removed. Figure 2.7 shows the spectra for the vowel /a/ with and without the presence of the near field microphone. There are some small differences (such as in the fundamental frequency), but in general the two spectra are similar in shape, and thus it is likely that the near field microphone is not confounding the far field pressure measurements.

Step Four: Estimating the vocal tract transfer function.

Given that both $T_1(f)$ and $R_n(f)$ are now both known quantities, the remaining task is to separate $VT(f)$ and $N(f)$, i.e., the remaining composite transfer function is $C(f)=VT(f)N(f)$. At this point, it is helpful to take advantage of the properties of these two different transfer functions. First, based on measurements of other body tissues, one would expect $N(f)$ to be slowly varying with frequency, with few, if any, sharp resonances (Wodicka, Lam, Bhargava & Sunkel, 1993). The vocal tract transfer function, on the other hand, can be effectively modeled as an all pole system (at least for the production of non-nasal vowels), with the location of the resonances and bandwidths of the resonances are dependent on the vowel being produced. Therefore, if the center frequencies and bandwidths of the poles of vocal tract transfer function can be estimated, it should be possible to deconvolve this function from $C(f)$.

One common algorithm used to estimate the vocal tract parameters is linear prediction. Linear prediction assumes that the vocal tract transfer function can be represented by an all-pole filter (Atal & Hanauer, 1971). Thus, $T(z)$ (the z-transform of the impulse response of the vocal tract), can be modeled as

$$T(z)=\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}} \tag{11}$$

where a_k is the kth linear predictor coefficient, and p is the number of coefficients (usually 12-14 coefficients are used, depending on the sampling rate used to digitize the speech). The predictor coefficients are determined by minimizing the mean-squared prediction error, $<E^2[n]>_{avg}$, which is the average difference between the actual signal (in this case, $c[n]$) and its predicted value, $\hat{c}[n]$. Since an estimate of the vocal tract transfer function, $T(z)$, is now available an inverse filter, G(z), can be designed, where

$$G(z) = \frac{1}{T(z)} = 1 - \sum_{k=1}^{p} a_k z^{-k} \ , \tag{12}$$

and by multiplying $C(f)$ by $G(f)$ (which can easily be obtained from $G(z)$), the frequency response of the neck wall, $N(f)$, can be determined.

However, since the linear prediction algorithm is applied to $c[n]$, the estimate of the vocal tract will be corrupted by the presence neck transfer function, and if $c[n]$ is inverse filtered, then properties of the neck transfer function could be removed as well. Thus, before performing the inverse filtering, it is important to modify the estimate of the vocal tract transfer function, $G(z)$, such that it only consists of the vocal tract resonances, and not any resonances that may be associated with $N(f)$. Two tools were used to make this modification. First, the average values of formant frequencies and bandwidths for different vowels have been published (Peterson & Barney, 1952) and can be used as criteria for deciding between formant and non-formant poles. However, because the laryngectomy patient's vocal tract has been altered, the formant frequencies tend to be shifted higher in frequency because the vocal tract has been truncated. For males, the first formant frequency can shift by up to 250 Hz, although a 100 Hz shift was more common. Female esophageal speakers typically present a shift of about 50 Hz (Sisty and Weinberg 1972). Moreover, since the vocal tract is effectively separated from the trachea and lungs, there is no subglottal coupling and thus the bandwidths will be narrower (Fujimura &

Lindqvist, 1971; House & Stevens, 1958). Therefore, some correction had to be made in the criteria used to separate formant poles from non-formant poles.

The second tool is the data collected from the continuous vowel transition task. If a spectrogram is made from the pressure signal at the microphone, the formant transitions from vowel to vowel can be tracked. This formant tracking gives an estimate of the formant frequencies and bandwidths for each vowel.

In addition, because the formant frequencies should change as the vowel changes (at least for the first three formants), any constant resonances that are present in the spectrogram, are most likely not vocal tract resonances and should not be a part of the vocal tract transfer function estimate. Thus, the spectrogram can indicate which resonances are part of the vocal tract transfer function and which are not. Figure 2.8 shows an example of a spectrogram of the pressure signal collected during the vowel transition task.

An important point to consider is the fact that the placement of the shaker head against the neck wall produces a sound source that is located somewhere other than the end of vocal tract tube. As a result, a small back cavity is formed in the vocal tract with its own resonances that act as zeros in the vocal tract transfer function. The end result is a downward tilt of the vocal tract transfer function, especially at higher frequencies. Because the distance from the shaker head to the back of the vocal tract cannot easily be measured, the effect of the back cavity cannot be accurately predicted.

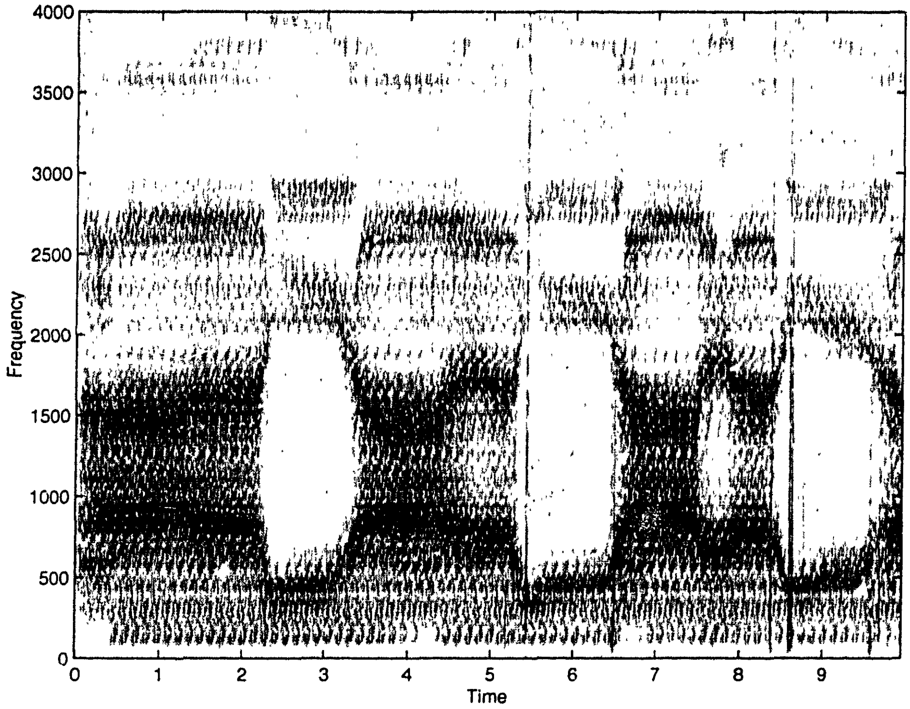


Figure 2.8. Spectrogram of the pressure signal collected during the vowel transition task. The vowel sequence is /a/-/ae/-/I/-/a/-/ae/ as can be seen by following the formant trajectories. The first two formants begin close to each other in the vowel /a/ and begin to spread apart as the subject moves to /ae/ and /I/. Those resonances that are not moving as the vowels change are most likely not formant resonances and should not be removed during the inverse filtering.

Once the LPC spectrum has been corrected such that it only includes vocal tract resonances, new linear prediction coefficients are computed and a new inverse filter, $\hat{G}(z)$, is formed from these new coefficients. The impulse response of the neck transfer function, $n[n]$ is determined by filtering $c[n]$ with $\hat{G}(z)$, and from $n[n]$, $N(f)$ can easily be computed.

This analysis algorithm was applied to the data collected from each trial so that the NTF computed from the data acquired with different vowels could be compared. Theoretically, the NTF should be independent of the vowel used in the experimental task, and this comparison can help verify the validity of the experiment and analysis. Nevertheless, it is possible that the NTF has a small resonance in the vicinity of a formant of a particular vowel, and that this resonance will be masked by the formant. But this NTF resonance may be "unmasked" by analyzing the data collected from a trial using another vowel.

2. Coherence and the LTI Assumption

For each open mouthed experimental task (other than the vowel transition task), the subject was asked to maintain a static vowel position for about 15 seconds. However, holding a vowel for this amount of time was extremely difficult and the subjects often changed the vowel that was being produced as the task progressed. If one were to compute the transfer function by analyzing over the entire task length then one would encounter two problems. First, if the subject is changing his/her vocal tract configuration, then the center frequencies of the formant resonances are moving over time, and if spectrum of the pressure signal is computed by averaging over the whole signal, the formant bandwidths may be wider than expected. Second, because the vocal tract transfer function is changing over time, the LTI assumption is violated and thus the procedure described above is no longer valid. Therefore, a method was needed to ensure that the signals used to compute the transfer function were produced from an LTI system. The coherence of two signals used to estimate a transfer function is a good measurement of how reliable the estimate of the transfer function is. If $x[n]$ is the input to a system and $y[n]$ is corresponding output, then the coherence between $x[n]$ and $y[n]$, $C_{xy}(f)$, is defined as

$$C_{xy}(f) = \frac{\left|S_{xy}(f)\right|^2}{S_{xx}(f) \cdot S_{yy}(f)} \tag{13}$$

where $S_{xy}(f)$ is the cross-spectral density of $x[n]$ and $y[n]$, and $S_{xx}(f)$ and $S_{yy}(f)$ are the power spectral densities of x[n] and y[n] respectively. The coherence ranges between 0 and 1, where 1 means that the transfer function estimate at that frequency is reliable. Any large deviations from

1 suggest that either the system in question is not linear and/or time-invariant, or that the measurements of $x[n]$ or $y[n]$ are noisy.

In the case of these measurements, the coherence between $p[n]$ and $v[n]$, $C_{vp}(f)$ is of interest. Since the system is being treated as two branches consisting of LTI systems, $C_{ap}(f)$ should be 1 for all frequencies. Therefore before computing the NTF for a trial, $C_{ap}(f)$ was computed and examined. In most instances, when the entire data set collected for one trial was used to compute the coherence, significant deviations from unity were found, especially in the frequency ranges corresponding to formant frequencies. These deviations suggested that the subjects were having difficulty maintaining a static vowel for an extended length of time. Therefore, only a section of data was chosen to be analyzed, such that its coherence was never below 0.8 for $100 < f < 3000$ Hz. . Figure 2.9 shows an example of an acceptable coherence spectrum. Beyond 3000 Hz, the coherence always significantly deviated from 1, no matter how small the section of data was. . This procedure increased the likelihood of estimating a linear time-invariant system.

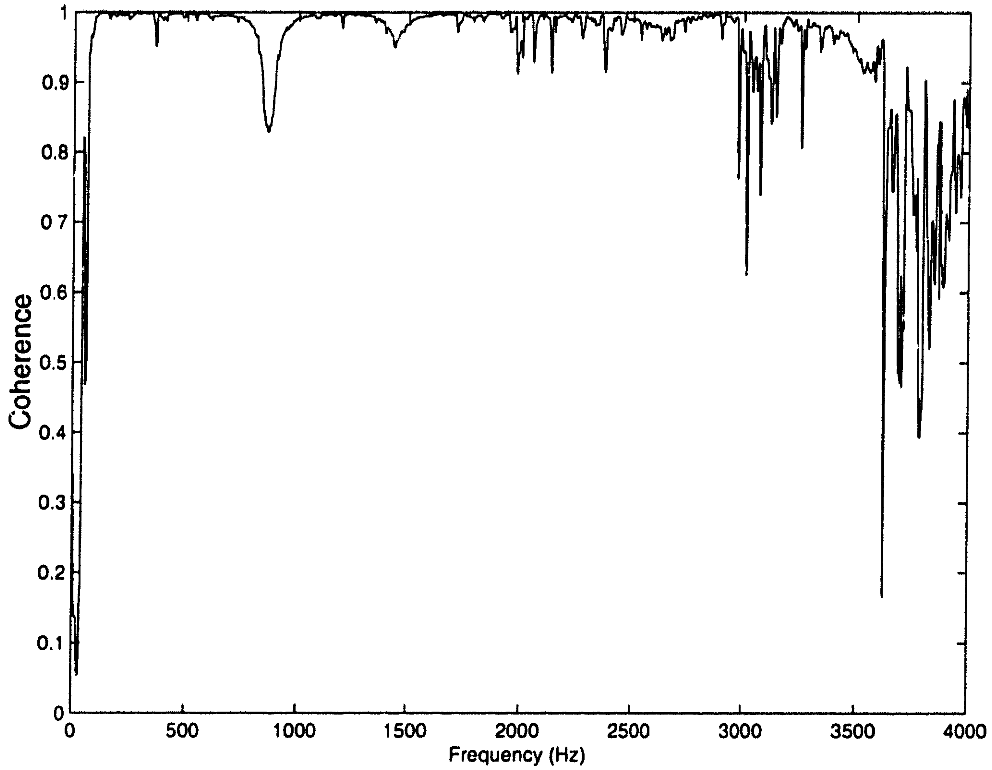


Figure 2.9. Acceptable coherence spectrum for the subject rsm. For frequencies between 100 and 3500 Hz, the coherence never drops below 0.8. The small dips in the coherence in this frequency range occur at the formant frequencies, in this case for the vowel /a/. These dips suggest that the vowel was not entirely static for the duration of the data segment, but 0.8 is not a significant deviation from 1. For frequencies above 3000, the coherence rapidly decreases, indicating that either the NTF is not LTI at these frequencies and/or there is noise corruption in the measurements at these frequencies.

3. Resampling the Data

As mentioned in the experimental procedure section, all of the data were sampled at 25 kHz, giving us data with analyzable spectra up to 12.5 kHz (although the 6 kHz anti-aliasing filters do reduce the analyzable bandwidth). However, this bandwidth extends beyond the frequency range of interest. The electrolarynx is responsible for producing voiced speech whose energy is concentrated below 4 kHz and thus attempting to measure the neck transfer function at higher frequencies is not necessary. Moreover, as the coherence spectrum in Figure 2.9 demonstrates, at these higher frequencies, either the LTI assumption no longer holds and/or the measurements are too noisy to accurately measure the NTF. Finally, using signals with larger bandwidths requires using larger order LPC filters to attempt to estimate the vocal tract transfer function. It is possible that there are some resonances unrelated to the vocal tract present at these higher frequencies and the LPC algorithm would attempt to try and fit these poles as well, making the estimation of the vocal tract poles less accurate. Early attempts to use full bandwidth data produced LPC fits that used multiple poles to represent the vocal tract resonances, thus making it more difficult to separate formant poles from non-formant poles. Given these reasons, the data were resampled from 25 kHz to 8 kHz using a linear phase optimal anti-aliasing FIR filter.

4. Choosing the LPC Order

The order used in the LPC algorithm plays a significant role in the accuracy of the estimation of the formants. Typically, an LPC filter of order 12 can accurately represent a speech segment sampled at 10 kHz (Atal and Hanauer, 1971). However, in this case, the data is sampled at 8 kHz and the speech segment is not a typical speech segment in that it is contains both the neck transfer function and the vocal tract transfer function. Therefore, the proper order to use remains unknown and some decision criterion must be employed. Ideally, the decision should be based on some objective criterion and eliminate the subjectiveness associated with an observer basing the decision on which LPC fit "looks" best. Therefore, for each trial, LPC coefficients were computed using orders 10 though 16 along with their corresponding error signals. The error signals were computed by inverting the LPC filter (i.e. producing an all-zero filter) and using the inverse filter to filter the corrected pressure signal (i.e. $s[n]$ in Figure 2.4). The LPC order that produced the error signal with the smallest energy was the order used in the final analysis. In most cases, this method chose an LPC filter of order of 16 as the best fit.

5. Minimum Signal Quality Criteria

Before the set of signals from each trial could be analyzed, it had to meet a set of minimum quality criteria. First, the spectrum of the measured open mouth pressure signal had to display at least two (preferably three) prominent formant peaks. If the pressure signal spectrum does not have enough formants then the LPC algorithm would be unable to estimate the vocal tract transfer function and thus the analysis algorithm would fail. Second, the computed coherence associated with a set of signals could not demonstrate large deviations from unity over significant part of the frequency range 100 Hz to 3000 Hz. If the coherence was significantly less than one, then either the systems that were measured were not LTI (i.e. the subject had difficulty maintaining the static vowel position) and/or the measurements were corrupted by additive environmental noise. Those signal sets that did not meet the minimum criteria were not analyzed. As a result, one female laryngectomee's entire data set was discarded since every trial could not meet the requisite criteria. This particular subject was unable to use the neck-type electrolarynx because it was unable to transmit enough sound through her neck to produce audible, intelligible speech. Not surprisingly the signals recorded at the lips were not of sufficient magnitude to accurately estimate the neck transfer function.

6. Statistical Analysis

To help simplify the comparison of the NTF across subjects, five features were extracted from the data that capture the spectral pattern of the NTF: Peak gain, Peak frequency, maximum attenuation, low frequency slope, and high frequency slope. The peak gain is the maximum of the magnitude of the NTF (in dB re 1 cm^2-s), and the peak frequency is the frequency at which the peak gain is located. Maximum attenuation is defined as the difference between the peak gain and the minimum gain. The low frequency slope is the slope of the magnitude of the NTF between 100 and 200 Hz, while the high frequency slope is defined as the slope of the magnitude of the NTF between 1000 and 2000 Hz.

The subjects were grouped by sex and by laryngeal status; laryngectomy or normal larynx. The mean and standard deviation of the five parameters were computed for each group. In addition, T-tests were performed to determine if there were any significant differences in average data values between groups. Because of the exploratory nature of the data collected in this study, a p value $\leq .05$ was considered statistically significant.

III. Experimental Results

A. General Description

The results reported in this section were obtained from 9 laryngectomy patients (5 male, 4 female) and 4 normal or laryngeal subjects (2 male, 2 female). Although there was significant inter-subject variation, the neck transfer function data did demonstrate some general trends. Figure 3.1 gives two examples of typical neck transfer functions, one from a laryngeal female subject and one from a male laryngectomized subject. In most cases the NTF peaked at around 100 Hz, (109 Hz ± 26 Hz), rolled off at a slope of -10.2 ± 3.8 dB/octave between 100 Hz and 200 Hz, flattened out between 300 Hz and 1000 Hz, and then decreased at a slope of -5.22 ± 0.75 dB/octave between 1000 Hz and 3000 Hz. The NTF of several subjects displayed one or two small peaks with center frequencies located between 200 and 500 Hz ($f_{c1}=314 \pm 100$ Hz, and $f_{c2}=364 \pm 40$ Hz, for the first and second peaks respectively).

As discussed in the previous section, five features, peak gain, peak frequency, low frequency slope, high frequency slope and maximum attenuation, were extracted from the measured neck transfer functions in order to make comparisons across different subject groups. These quantities are useful for simplifying the comparison of the NTF across subjects and neck position, and yet capture the most salient features of the NTF. T-tests were performed on means and standard deviations for each of these features to help verify any trends that were found in the NTF spectra of different subject sub-groups. Any conclusions, however, that can be drawn from the statistical analysis is limited by the small number of subjects in each group (in some cases as small as N=2).

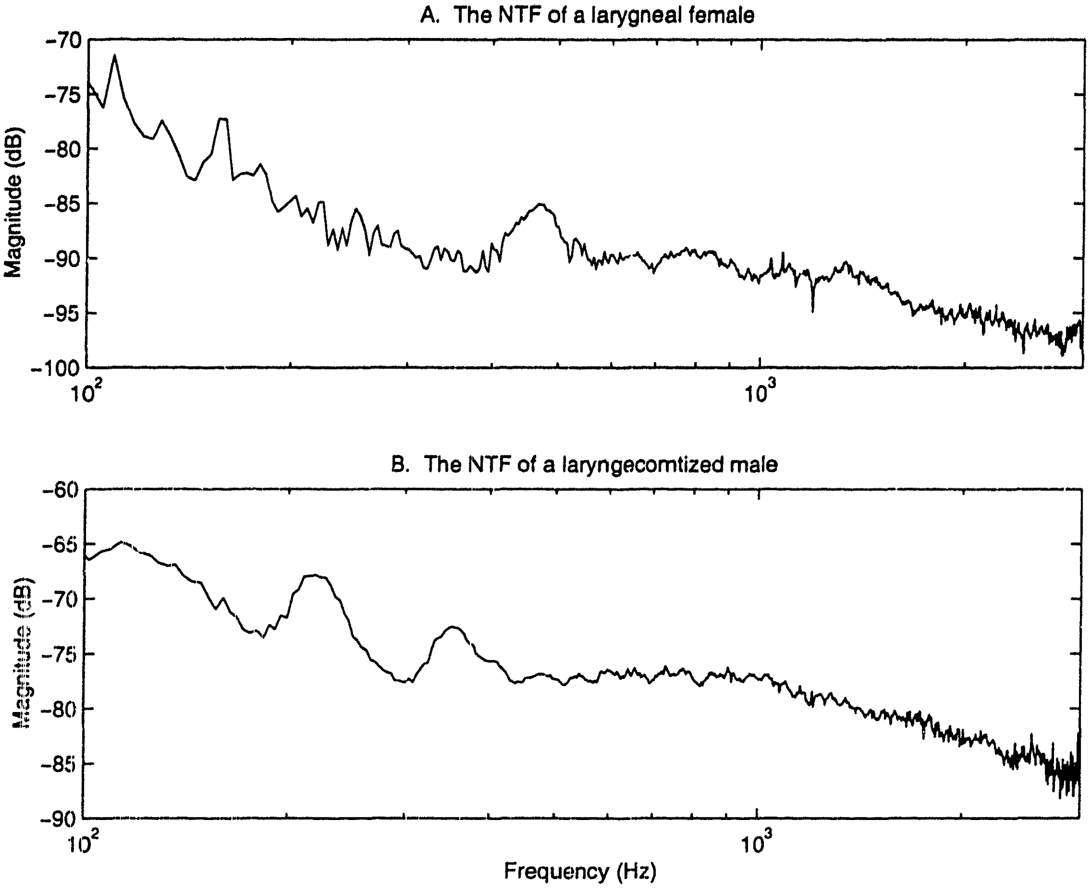


Figure 3.1. Examples of two neck transfer functions. A. The neck transfer function of a laryngeal female. This plot is the average of the transfer functions computed using three trials for each of the three vowels (i.e. 9 trials total). The magnitude peaks at 109.4 and decreases at –10 dB/octave until 300 Hz, where it flattens out until 1000 Hz . This NTF presents a single small peak at 460 Hz. B. The neck transfer function at position 1 for a laryngectomized male. This NTF peaks at 101.6 Hz, decreases at a slope of –10 dB/octave until 300 Hz, then flattens out until 1000 Hz where it then decreases at a slope of –4 dB/octave. This particular NTF displays 2 small peaks at 225 Hz and 360 Hz. Note that the magnitude scale is in dB re 1 cm^2-s.

B. Reproducibility of Results

The experimental procedure was designed to include two separate methods to check the reliability of the measurements. First, at each position of the neck, measurements were repeated three times per vowel. Ideally, for a single position on a single subject, the NTFs computed using the same vowel should be exactly the same. However, as in most experiments some trial to trial variation is expected. Repeating the NTF measurements and calculations with the same vowel can provide an idea of the reproducibility of the results. Figure 3.2 shows an example of the "intra-vowel" variation for a male laryngectomy subject. In general, the "intra-vowel" repeatability was quite good; the average (over frequency) standard deviation was 2.05±0.81 dB. There was some variation in the reliability from vowel to vowel as the vowel /a/ demonstrated more variability than the other vowels: the average standard deviation for /a/ was 2.41 dB, while for /ae/ and /I/ it was 1.71 dB and 1.99 dB respec' 'vely.

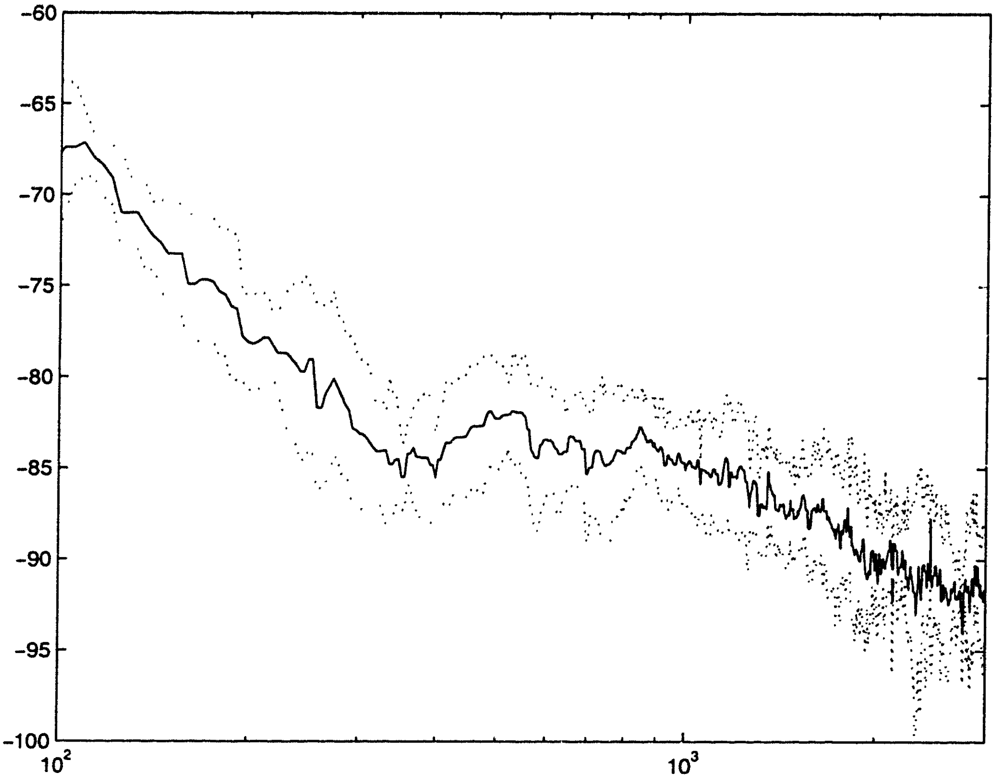


Figure 3.2. Average of the NTF estimation for a male laryngectomy using the vowel /a/. The solid line represents the average transfer function while the dotted lines represent the standard deviation at each frequency. These values were computed by averaging over the three trials using the vowel /a/ at one neck position. The standard deviation ranged from 0.25 dB at 836 Hz to 4.5 dB at 121 Hz.

The second type of variation that can be examined is "inter-vowel" variation. Since the NTF is measuring the sound transmission properties of the neck tissue, it should be independent of the vocal tract configuration. Therefore, a good test of the experimental procedure and analysis algorithm is to examine the variability in the NTF estimation when using different vowels, i.e. intra-vowel variation. Figure 3.3 shows a typical example of the inter-vowel variability for a single subject. Although the NTFs computed using the different vowels were not exactly alike, the inter-vowel variability was small. Averaged over the frequency range 100 Hz – 3000 Hz, the standard deviation of the NTF ranged from 1.4 to 3.6 dB with a mean value of 2.5 dB. The small amount of variability from vowel to vowel suggests that the NTF is indeed independent of vocal tract configuration (as expected) and helps validate the experimental procedure and analysis algorithm.

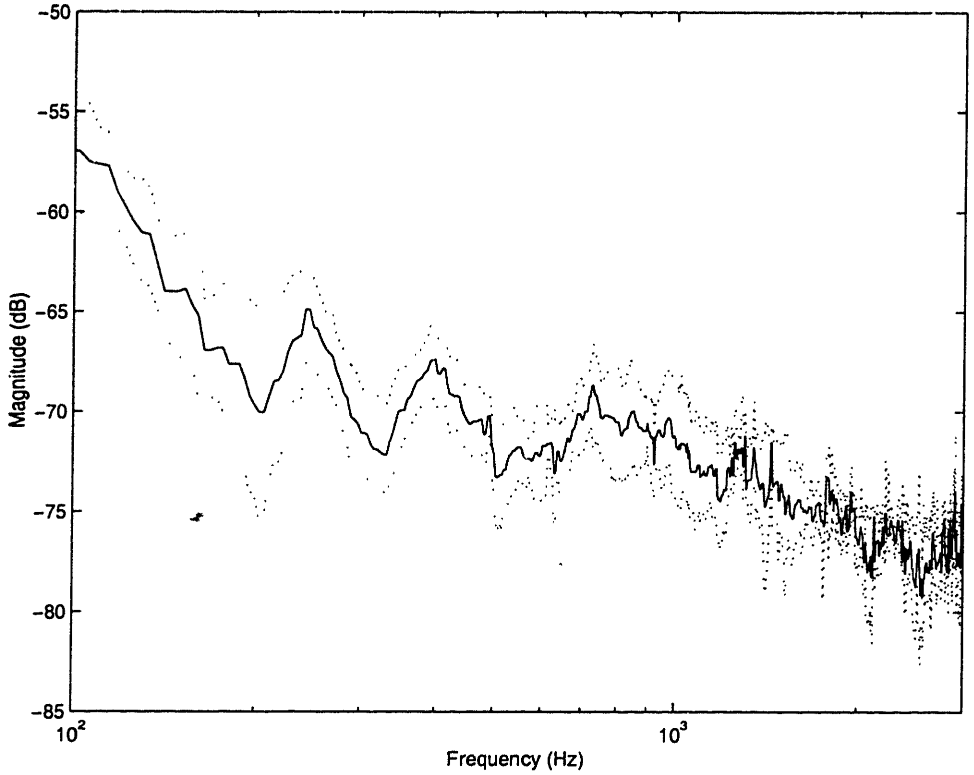


Figure 3.3. The average (solid line) and standard deviation (dashed lines) of the NTF at a single position on the neck of a female laryngectomy patient. The average NTF for a single position was computed by averaging the all of the trials for a single position. This particular case produced a mean standard deviation of 1.23 dB with a maximum of 5.35 dB at 203 Hz and a minimum of 0.17 dB at 2324 Hz.

C. Variations in the NTF

One of the goals of this research was to determine how the NTF changed (if at all) from subject to subject. More specifically, it was desired to learn how the NTF differed between males and females, between laryngectomees and normals, and between different locations along the neck.

1. Neck location

As discussed in section II-B, measurements were made on three locations on the neck except in certain cases where it was not possible. From these measurements, it was determined that the NTF dependence on neck location was highly subject dependent. The type of location dependence could be classified into three groups. The first group showed little or no dependence on neck location. This group consisted of both laryngeal male subjects and one laryngectomized subject. It is not too surprising to find that the NTF of the laryngeal male subjects was relatively independent of location because they have not undergone any surgical procedures. The one laryngectomy subject that fits in this category is somewhat unexpected since laryngectomy operations tend to produce asymmetries in the neck. Nevertheless it is possible that the tissue characteristics are more or less uniform in the spots where the measurements were made for this one subject. Figure 3.4 gives examples of location independent NTFs for the three different positions on the neck of one laryngeal male and one laryngectomized male.

The second group demonstrated a dependence such that the shape of the NTF remained constant, but the gain changed. This group included both laryngeal female subjects, 2 laryngectomized males and 3 laryngectomized females. Figure 3.5 gives two examples of NTFs that fit in this group (one laryngeal and one laryngectomized female). The location dependence found in the laryngeal female subjects was unanticipated, especially given the independence shown in the laryngeal males. In both laryngeal females, the NTF at Position 2 demonstrated a notably smaller gain for all frequencies (See Figure 3.5 for an example). Position 2 was located 2 cm below Position 1 (and Position 3) so one might suppose that the change in gain is due to some increased attenuation associated with the increased distance from the shaker driving point to the microphone.

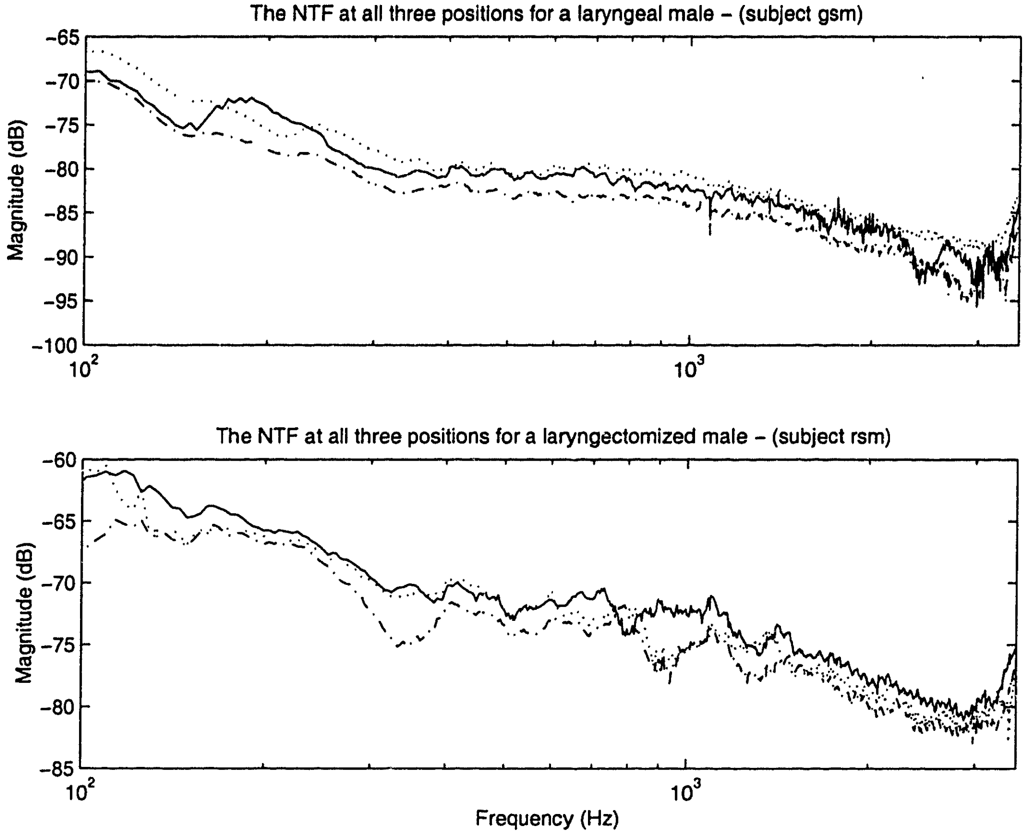


Figure 3.4. Example of the estimated NTF at Position 1 (solid), Position 2 (dotted), and Position 3 (dot-dashed) of a laryngeal male and a laryngectomized male. Although there are some small variations from position to position, the main features of the spectral pattern remain more or less constant. The amount of variability from position to position was similar to the intra-vowel and inter-vowel variability.

However, in other subjects in this group (for example, the laryngectomized female in Figure 3.5), the NTF at Position 2 did not present the smallest overall gain, so the change in distance cannot entirely explain this phenomenon. One possible explanation for the difference between laryngeal males and females is that for the female laryngeal speakers, Positions 1 and 3 were located above the larynx, while Position 2 was located on the larynx (thyroid cartilage) and thus representing a significant change in the type of tissue, possibly producing the change in gain.

The male laryngeal speakers tended to have larger necks and therefore a change in location of 2 cm may not have caused as much change in the underlying tissue properties between Positions 1 and 2 as in the laryngeal females, and thus the gain did not change in these subjects. The location dependence of the NTF in laryngectomized subjects was expected to occur since the tissue properties of the necks of laryngectomees often appear to change with location. In many cases, certain parts of the neck are quite rigid due to scarring and the effects of radiation therapy, while other parts are soft and pliable. Thus it is possible that the changes in the NTFs are reflecting these differences.

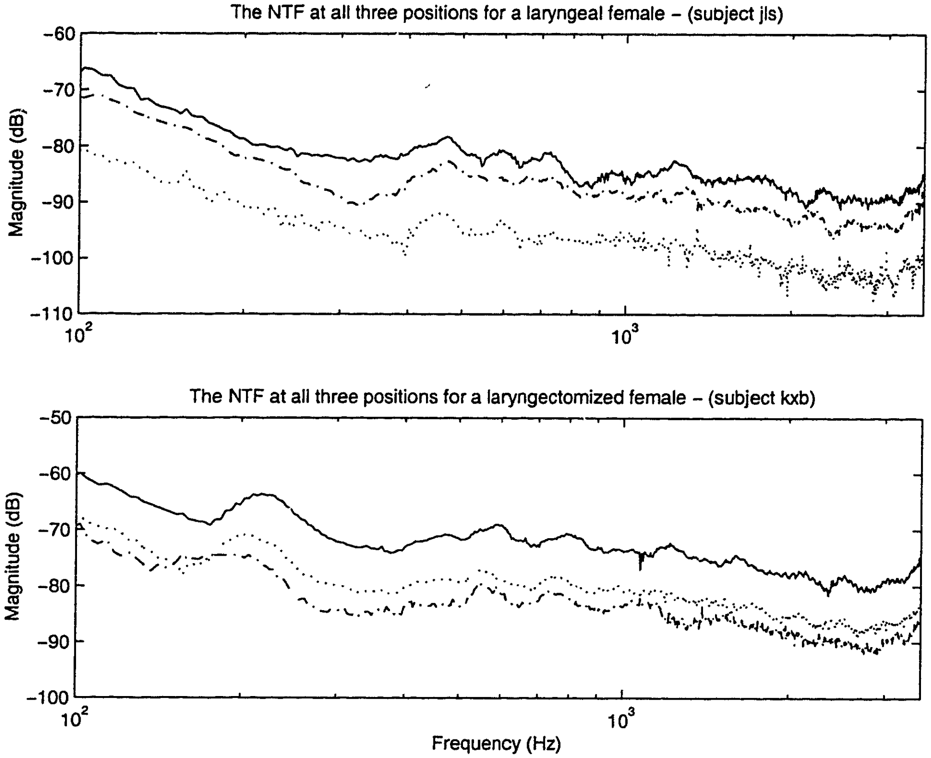


Figure 3.5. Example of the estimated NTF at Position 1 (solid), Position 2 (dotted), and Position 3 (dot-dashed) of a laryngeal female and a laryngectomized female. In both cases, while there's a noticeable change in gain in at least one position, the overall shape of the NTF does not change. In the case of the laryngeal female, Position 2 demonstrates the smallest overall gain, while in the laryngectomized female the NTF at Position 3 has the smallest gain.

The third group presented NTFs whose gain and shape were dependent on neck location. The only subjects in this group were laryngectomees: 2 male and 1 female. Figure 3.6 shows the NTFs at all positions for these three subjects. Due to the nature of subject *lxh*'s neck, the measured signals at Position 2 were given low ratings (as per the rating system discussing in Section 2) and thus an accurate estimation of the NTF could not be made at this position.

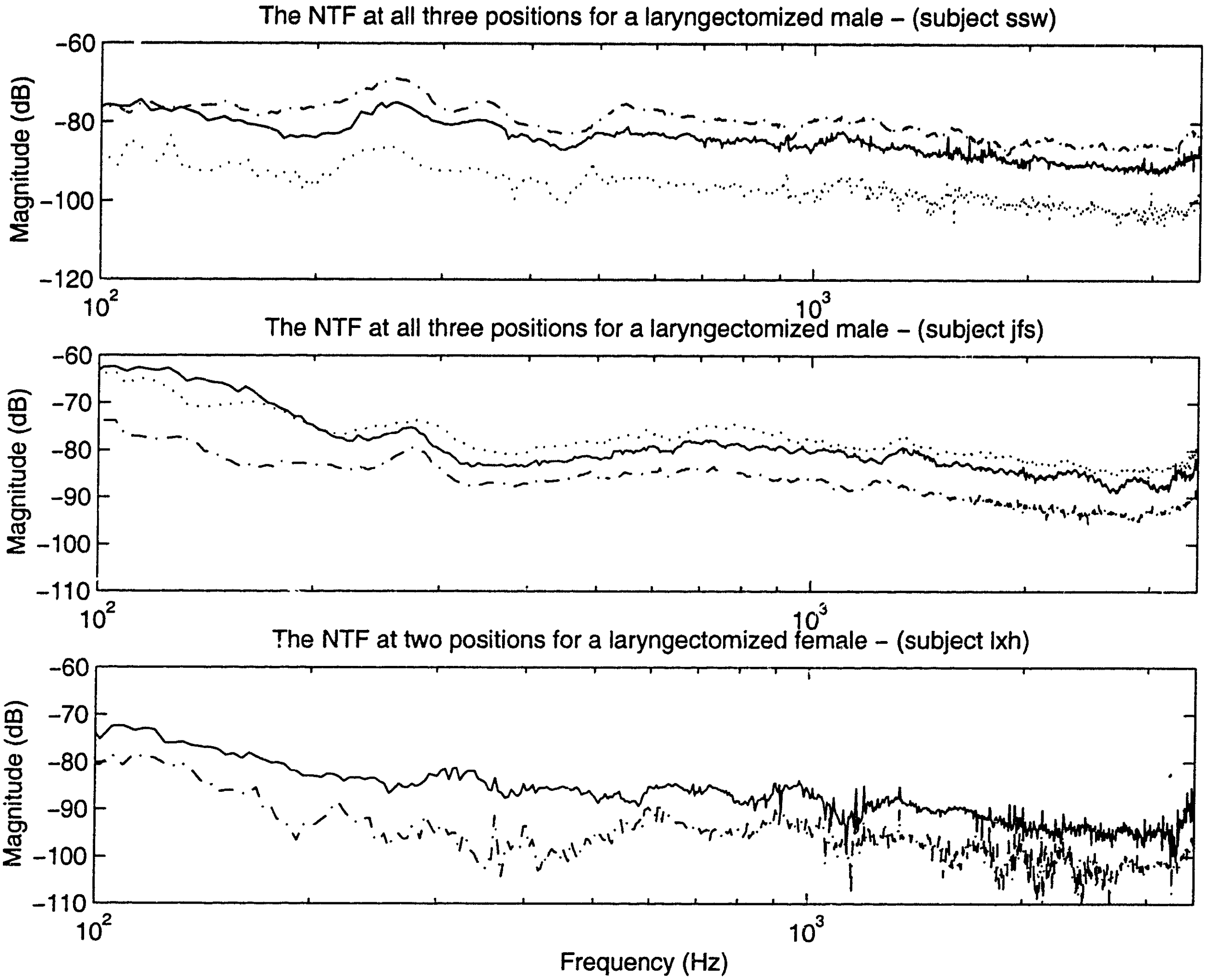


Figure 3.6. The NTFs at all three positions (Position 1 (solid), Position 2 (dotted), and Position 3 (dot-dashed)) for the three laryngectomy patients who had location dependent NTFs. In the cases of subject ssw, the NTF changes from having a negative slope between 100 Hz and 200 Hz in Positions 1 and 2 to having a slightly positive slope in Position 3. For subject jfs, the NTFs of Positions 1 and 2 decrease at –12 dB/octave at low frequencies while in Position 3 the slope starts at –8 dB/octave and then increases to 0 slope between 150. and 250 Hz. Subject lxh shows notable differences in the NTFs between 300 Hz and 600 Hz. Note that an estimate of the NTF at Position 2 could not be made in this subject due to the poor quality of the signals obtained.

An inspection of Figure 3.6 reveals that for all three subjects, the NTF of Position 3, the transfer function at Position 3 always differed from those at the other positions. Positions 1 and 2 were only 2 cm apart and both located on the same side of the neck, so it is reasonable to expect the neck tissue to remain uniform over such an area. Even for those subjects who fit into the second category that produced an NTF at Position 2 with a smaller gain, the shape of the NTF remained constant, indicating that the tissue properties probably did not change (possibly there was just more tissue at this location). Position 3, however was located on the opposite side of the neck, often in a position where the subjects claimed the electrolarynx could not generate enough sound to produce audible speech. Thus, it is not surprising to learn that the NTF at this position differs from the NTFs at the other positions.

2. Laryngectomee vs. Laryngeal

Because of the changes in the neck tissue that accompany a laryngectomy and radiation therapy, it was expected to find some differences between the NTFs of the laryngectomized subjects and those of the non-laryngectomized subjects. However, if the NTFs are averaged over each group, i.e. all laryngectomees and all non-laryngectomees, little difference is found between them. Figure 3.7 displays the average NTF with its standard deviation for each group. The two NTFs look very similar in shape, with the peak at 100 Hz, a decrease at about –10 dB/octave until 350 Hz, at which it remains constant until 1000 Hz where it then decreases at –5 dB/octave up to 3000 Hz. Although the average NTF of the laryngectomized subjects has a gain that is 3.2 dB greater on average than that of the NTF of the laryngeal subjects, this difference falls with in the standard deviations of both NTFs. Furthermore, the statistical tests performed on the values of the five features further support the claim that there is little difference between the NTFs of these two groups. Table 3-1 presents the mean values for the five features of the

NTFs for both groups. The raw data for these features can be found in Appendix B. To compute this data, the value of each feature was computed for each transfer function of each subject and then the mean and standard deviation were computed using these values.

Feature	Laryngeal	Laryngectomee	p-value
Peak Gain (dB re 1 cm^2-s)	-71.1±6.3	-67.6±6.1	0.137
Peak Frequency (Hz)	105.5 ±5.2	111.2 ±30.6	0.365
Maximum Attenuation (dB re 1 cm^2-s)	24.9±2.2	21.8±9.4	0.553
Low Freq. Slope (dB/octave)	-9.9±5.4	-10.3±4.4	0.716
High Freq. Slope (dB/octave)	-5.4±0.7	-5.2±0.8	0.123

Table 3.1. Average values of features of the NTF for the laryngectomee and non-laryngectomee group. In terms of these parameters, the two groups demonstrate very little difference.

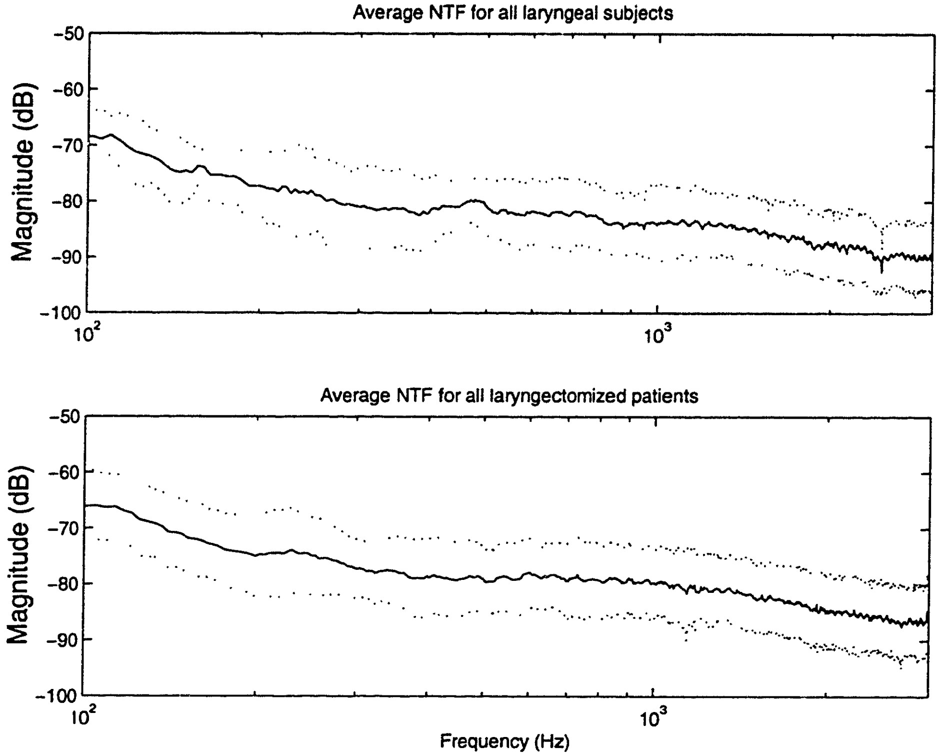


Figure 3.7. The average NTF (solid line) and standard deviation (dashed line) for the laryngectomee and laryngeal groups. There is little difference between the two transfer functions as further verified by the statistical analysis.

Based on the mean values in Table 3.1, it is all five NTF features appeared to be quite similar for the laryngeal vs. laryngectomee groups. A t-test performed on this data found that there was no significant difference between the two groups with regard to these parameters. The data was further analyzed by separating the laryngectomee and laryngeal groups by sex. The average transfer functions for these subdivisions are displayed in Figures 3.8a and 3.8b and the associated summary statistics are displayed in Table 3.2.

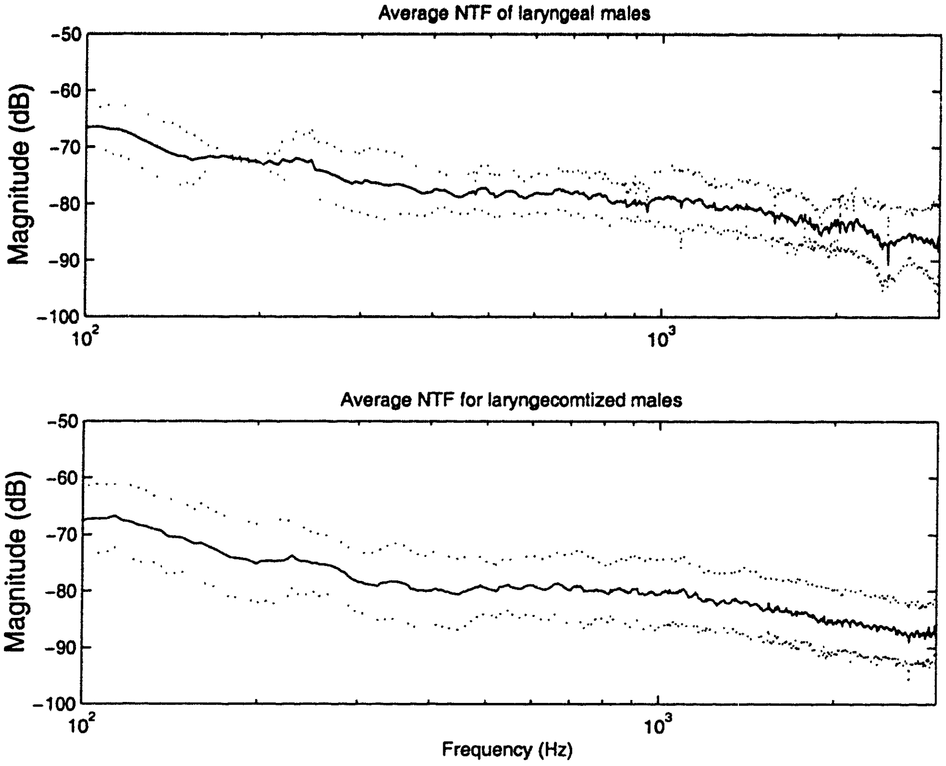


Figure 3.8a. The average (solid line) NTFs for laryngeal and laryngectomized males with their corresponding standard deviations (dashed lines). The two transfer functions are very similar as confirmed by the statistical data.

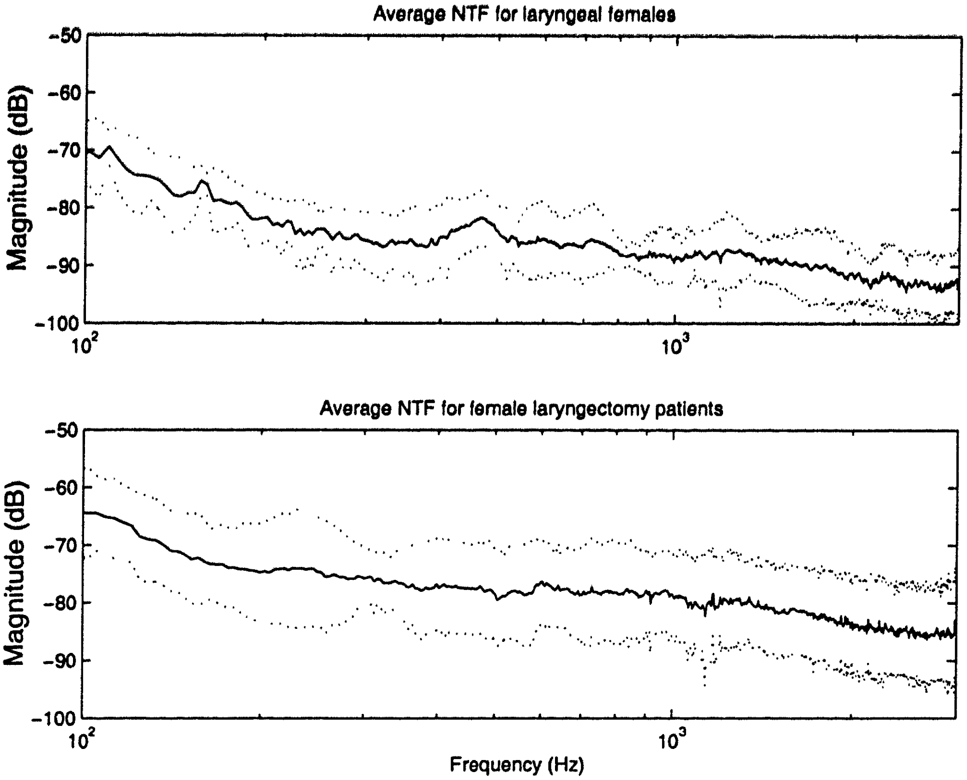


Figure 3.8b. The average (solid line) NTFs for laryngeal and laryngectomized females with their corresponding standard deviations (dashed lines). Although the overall shape of the transfer functions are similar (although the laryngeal NTF does show a small peak at 471 Hz), the NTF of the laryngectomized females is on average 8 dB greater than that of the laryngeal females.

	Male			Female		
Feature	**Laryngectomee**	**Laryngeal**	**P-value**	**Laryngectomee**	**Laryngeal**	**P-value**
Peak Gain (dB re 1 cm^2-s)	-68.3±6.2	-66.9±2.5	0.497	-66.7±6.0	-74.6±6.4	**0.034**
Peak Frequency (Hz)	117.5±39.5	104.7±6.9	0.249	102.7±3.9	106.1±3.8	0.104
Maximum Attenuation (dB re 1 cm^2-s)	20.2 ±12.0	23.1±1.5	0.382	24.0±2.3	26.4±1.5	**0.019**
Low Freq. Slope (dB/octave)	-8.5±4.3	-9.2±1.8	0.634	-12.6±3.4	-10.5±1.2	0.079
High Freq. Slope (dB/octave)	-5.1±0.9	-5.4±0.5	0.340	-5.27±0.6	-5.33±0.8	0.879

Table 3.2. Average values of features of the NTF for laryngectomees and laryngeals after being separated by sex. Although there are no significant differences found between the laryngeal and laryngectomized males, significant differences do appear between the peak gain and maximum attenuation in the female laryngectomees and laryngeals. There also appears to be a close to significant difference between the low frequency slopes of the two female groups. P values ≤ 0.05 are indicated in bold type.

This further separation of the subjects into subgroups reveals an interesting dichotomy. According to the average NTFs and the corresponding statistics, there are no significant differences between the NTFs of laryngeal and laryngectomized males. However, this is not the case for the female subjects. As can be seen in Figure 3.8b, the average NTF of the laryngectomized females has a gain that is on average 8.1 ± 1.1 dB greater than that of the NTF of laryngeal females. A t-test performed on the peak gains further supports that there is a significant difference ($p=0.034$) between the gains of the NTFs. Moreover, not only do the NTFs differ in overall magnitude, but there is also somewhat of a distinction in the spectral shape as indicated by the statistically significant difference in maximum attenuation ($p=0.019$) and a close to significant difference ($p=0.079$) in the low frequency slopes. However, it should be noted that while the differences in the two parameters that suggest distinct NTF shapes are statistically (or close to) significant, the quantitative differences are not large (2.4 dB in maximum attenuation, and 2.1 dB/octave in low frequency slope) especially when compared to the difference in peak (and overall) gains.

These results are somewhat perplexing because there is no obvious explanation as to why the NTFs of laryngeal and laryngectomized female differed, but those of the laryngeal and laryngectomized males didn't. One possible explanation is that the laryngeal female speakers were not as effective as their male counterparts in sealing off the subglottic system from the vocal tract when holding the vocal tract in the desired configuration. If the subglottic structures were not completely decoupled then they would have introduced losses into the system and attenuated the pressure signal measured at the microphone, thus producing an NTF with a smaller gain. However, if the subglottic system were coupled to the vocal tract, one would expect to find extra poles and zeros in the acoustic spectra (Klatt and Klatt 1990), but no such extra poles and zeros were found. Therefore, this hypothesis is not supported by the acoustic data.

The difference between the two mean NTFs could be related to the relative sizes of the male and female vocal tracts. On average, the female vocal tract is 15% shorter that that of males and most of this difference is in the pharynx (Klatt and Klatt 1990). Therefore, because of the smaller female vocal tract, it is possible that all three measurement positions on the female necks were located over some part of the thyroid cartilage while for the male subjects, the measurements were only made over soft tissue. Thus the difference in underlying tissue may be responsible for the different NTF data.

Finally, it is important to recall that measurements were only made on 2 normal female subjects and thus these differences in the normal female NTFs could be the result of sampling error associated with a small sampling size. Further testing of laryngeal female subjects would be required in order to solidify any conclusions drawn from these data.

3. Male vs. Female

Although there is no intuitive reason for there to be a difference between the NTFs of males and females (especially between laryngectomized males and females), early pilot work using 6 subjects suggested otherwise. This new set of experiments, however, indicates that at least for all (laryngeal and laryngectomized) males and females, the NTF does not have a sex dependence. Figure 3.9 shows the average NTF for the male and female subjects with their respective standard deviations and Table 3.4 has the corresponding statistical data. These data suggest that the average NTFs for males and females are quite similar. Although the male NTF tends to have a larger gain across all frequencies (except near 100 Hz), this difference is minimal (mean of 1.4 dB). Furthermore, according to the data in Table 3.4, the only statistically significant difference between the two groups is in the low frequency slope. The variation in slopes can be seen in Figure 3.9 but again, although the difference is statistically significant, the absolute difference is small (-3.2 dB/octave). If these two NTFs were used to help custom design different electrolarynx driving signals, the differences in resulting speech would probably not be noticeable.

Feature	**Male**	**Female**	**p-value**
Peak Gain (dB re 1 cm^2-s)	-67.9±5.5	-69.4±7.1	0.478
Peak Frequency (Hz)	114.3 ±34.6	103.9±4.2	0.197
Maximum Attenuation (dB re 1 cm^2-s)	20.9±10.4	24.8±2.4	0.120
Low Freq. Slope (dB/octave)	-8.7±3.8	-11.9±2.9	**0.007**
High Freq. Slope (dB/octave)	-5.2±0.8	-5.3±0.7	0.562

Table 3.4. Average values of features of the NTF for males and females. There appears to be statistically significant differences in terms of the peak gain, effective attenuation and the two resonances. P values ≤0.05 are indicated in bold type.

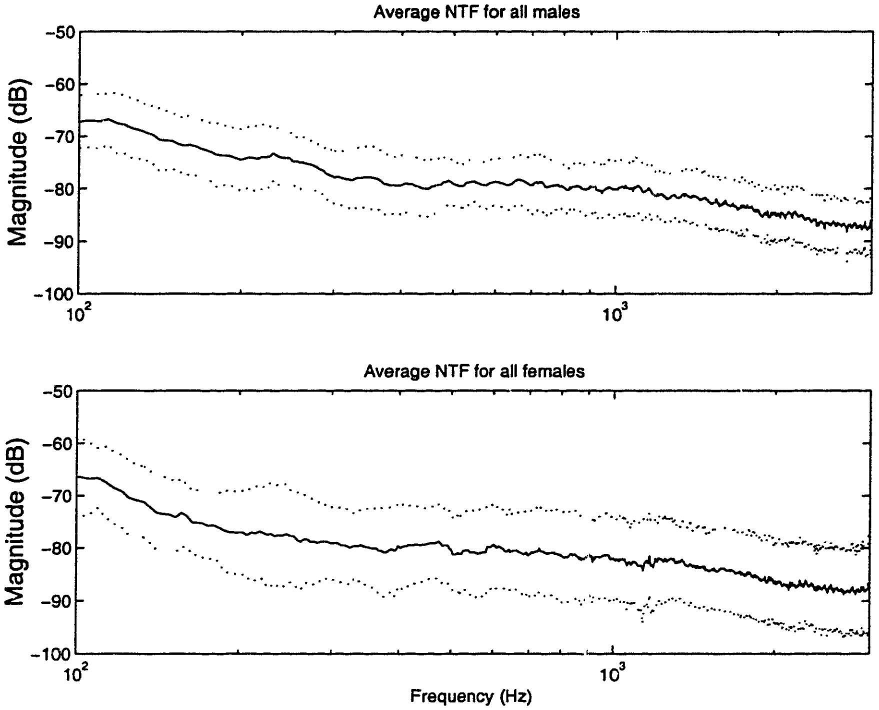


Figure 3.9 The average NTF (solid line) and standard deviation (dashed line) for all male subjects and all female subjects. The male NTF has a lower peak gain, smaller maximum attenuation, and smaller effective attenuation than does the female NTF.

Finding that the properties of the NTF do not depend on the sex of the subject is not unexpected, especially in the case of laryngectomized subjects. While it is conceivable that the NTFs of laryngeal males and females would differ due to sex-based anatomical-morphological variations, one would anticipate that for laryngectomees, the post-laryngectomy variability related to differences in surgical procedures would supersede any variability associated with sex. It would appear that the data in Figure 3.9 and Table 3.4 support these expectations. Yet, the data in Figures 3.8a and 3.8b and Table 3.3 seem to indicate that there indeed may be a sex related difference, since for males, the NTFs of laryngeal and laryngectomized subjects were very similar, but for females, the two groups had noticeable difference NTFs. To further explore this issue, the male and female subjects were divided into laryngectomee and non-laryngectomee categories and the corresponding summary statistics for the NTF features were computed for these groups and are displayed in Table 3.5. By dividing the subjects into these four groups, any variations found in the NTF should be able to be attributed to either the state of a larynx and/or to sex related differences. To illustrate the differences between the males and females of each group, the average NTFs displayed in Figures 3.8a and 3.8 were compared across sex, as shown in Figures 3.10a and 3.10b. Again, t-tests were performed to assess whether or not there were significant differences between males and females.

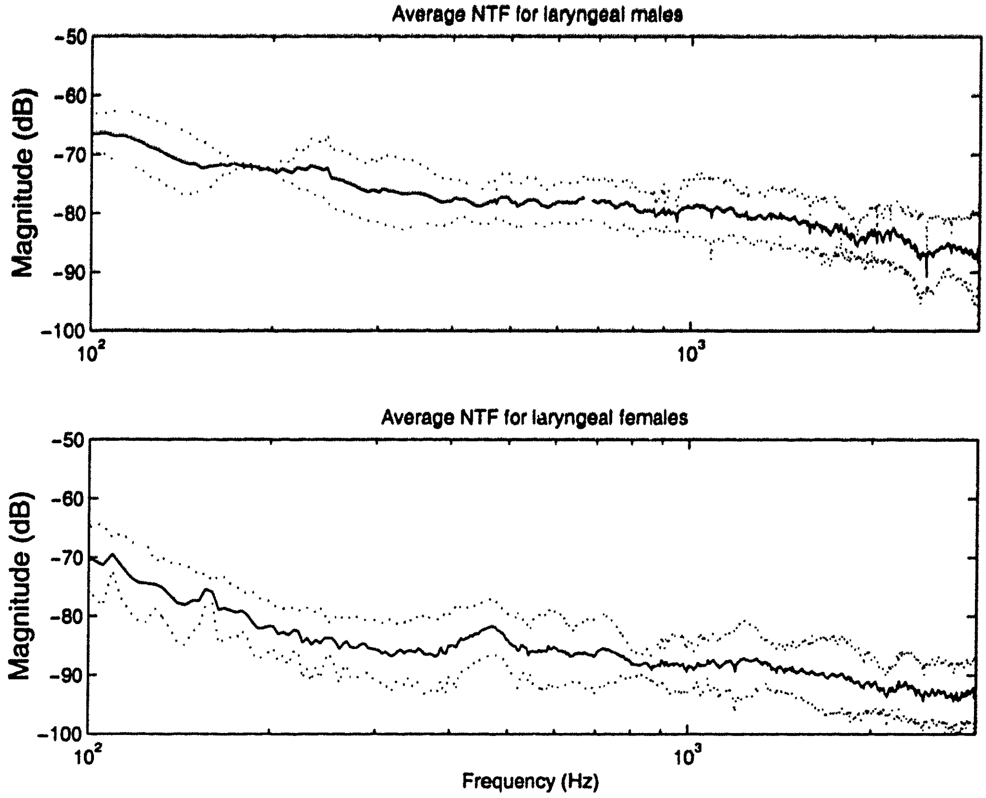


Figure 3.10a. The average (solid line) NTFs for laryngeal males and females with their corresponding standard deviations. There are some obvious differences between the two transfer functions in terms of peak gain, overall gain, low frequency slope, and maximum attenuation. Furthermore, the average female NTF presents a small peak at 471 Hz, while the male NTF shows no such peak.

. Again these statistics were computed to help validate the claims made based on the average NTFs shown in Figures 3.10a and 3.10b, but any conclusions that can be drawn are limited by limited number (4) of laryngeal subjects that were used.

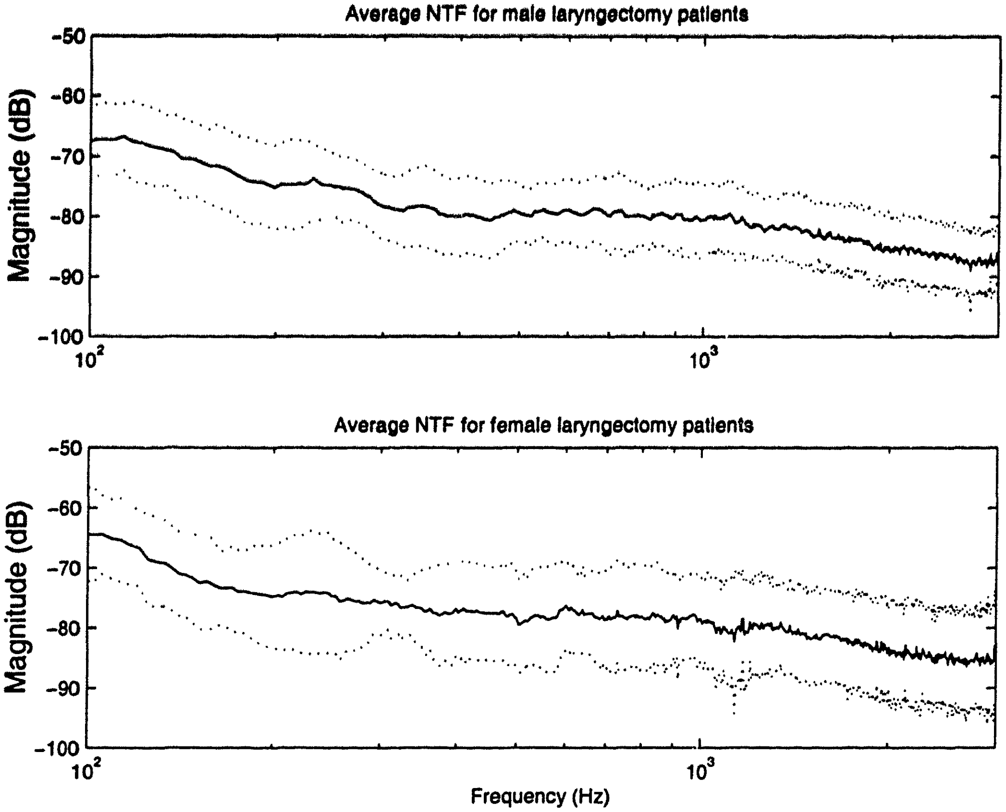


Figure 3.10b. The average (solid line) NTFs for laryngectomized males and females with their corresponding standard deviations (dashed lines). Although there is some difference in the low frequency slopes (-10.5 dB/octave for females vs. –8.3 dB/octave for males) the overall transfer functions are quite similar.

Comparing the NTFs in this fashion reveals that while there was only a small difference between the NTFs of laryngectomized males and females (namely the low frequency slope), the NTFs of laryngeal males and females are distinct in several ways. Figure 3.10a shows that the mean NTF of the laryngeal males has a larger magnitude for all frequencies (mean of 7.6±1.4 dB greater over all frequencies) and a more gradual low frequency slope (-7 dB/octave vs. –12 dB/octave). These descriptive-based differences are validated by the statistical data in Table 3.5 which indicates that the two mean NTFs are significantly different in three of five categories.

	Laryngectomees			Laryngeals		
Feature	**Male**	**Female**	**P-value**	**Male**	**Female**	**P-value**
Peak Gain (dB re 1 cm^2-s))	-68.3±6.2	-66.7±6.0	0.519	-66.9±2.5	-74.6±6.5	0.034
Peak Frequency (Hz)	117.5±39.5	102.7±3.9	0.171	104.7±6.9	106.1±3.8	0.696
Maximum Attenuation (dB re 1 cm^2-s)	20.2 ±12.0	24.0±2.3	0.258	23.1±1.5	26.4±1.5	0.005
Low Freq. Slope (dB/octave)	-8.5±4.3	-12.6±3.3	0.012	-8.4±0.9	-10.5±1.2	0.010
High Freq. Slope (dB/octave)	-5.1±0.9	-5.2±0.6	0.499	-5.4±0.5	-5.3±0.8	0.876

Table 3.5. Average values of features of the NTF for males and females separated by whether or not they have been laryngectomized or not. Although there is evidence of a sex dependence in both subject groups, the difference is noticeably greater in the non-laryngectomee group.

It was not expected for the average NTFs of the laryngectomized males and females to be notably different. A total laryngectomy involves the removal of the entire larynx and its surrounding tissues and theoretically should eliminate most anatomical differences that exist between laryngeal males and females. Any anatomical differences that do exist between laryngectomees are most likely due to differences in surgical procedures. Therefore it is just as likely for a male and female laryngectomee to have similar (or different) neck tissue properties as is for two female laryngectomees to have similar (or different) neck tissue properties. Since there are no systematic differences in the neck tissue properties with respect to sex, when comparing the average NTFs of both groups, there is little difference between them.

The situation is different for the laryngeal subjects, as reflected by the different NTFs. Because the larynges remain intact any anatomical differences associated with sex (e.g. size of thyroid cartilage) could affect the NTF and thus cause the differences in the average NTF for each group. However, because the number of laryngeal subjects was small, the conclusions that can be made from these data are limited.

4. The Effect of Changes in DC Force

Perhaps the most difficult variable to control in the experimental procedure was the amount of force each subject used to hold the shaker against his/her neck. While the DC Force channel from the impedance head was carefully monitored during the experiment, maintaining a steady state DC Force against the neck was problematic. Therefore, it was desired to determine what effect, if any, changing the amount of DC Force applied to the neck had on the neck transfer function. Step 9 in the experimental protocol (described in Section II-B) was designed to accomplish this task. Each successive trial in Step 9 involved increasing the DC force applied to the neck, usually in increments of about 0.5 – 2 Newtons. It was difficult to specify a certain force increment for each trial because it was difficult for the subjects to change the DC force with such precision. Therefore any increase in force that caused a noticeable increase in the voltage from the impedance head output was deemed acceptable.

In almost all cases, the changes in DC Force did not significantly change the overall shape the NTF. Figure 3.11 shows an example of the results of the changes in DC Force. While there are some slight differences in the three spectra, especially between the one estimated for the light force versus the other two, the variations in the spectral features were no greater than those seen from trial to trial. For this particular subject, if the three spectra are averaged together, the maximum standard deviation was 7.2 dB (at 100 Hz) and the mean standard deviation over all frequencies was 1.64 dB, well with in the range of the trial to trial variability. The fact that there are some changes in the NTF when the DC force is changed raised the question of whether or not the trial to trial variability discussed in Section III-B was due, at least in part, to the subjects changing the DC Force from trial to trial. A further investigation of the DC Force variability from trial to trial became necessary to address this question.

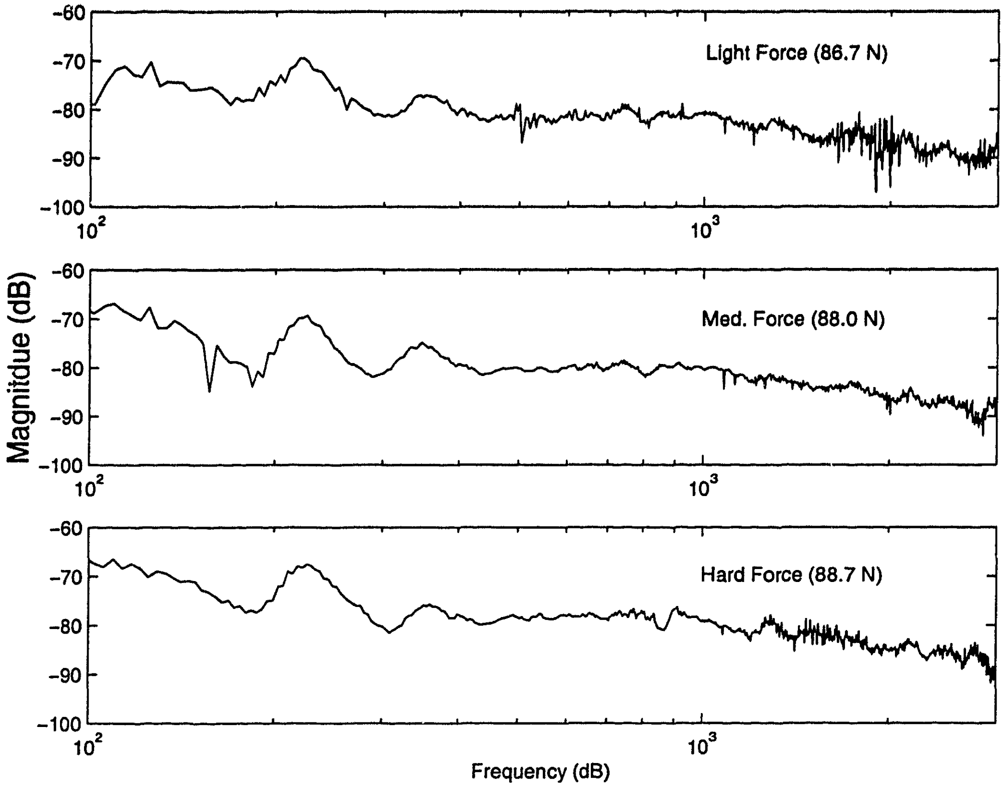


Figure 3.11. The NTF for different applied DC forces for a male laryngectomized subject. These plots show that while there is some change in the NTF (especially near 100 Hz), these differences are on the same order as the trial to trial differences and it appears likely that at least some part of the variability is due to the trial to trial variability.

The subjects were able to hold the shaker with a remarkably consistent force. For almost every subject, the trial to trial change in DC Force was no more than ±0.22 N. However the minimum change in DC Force for the DC Force task was 0.5 N and most changes were greater than 1 N. In the case of the subject in Figure 3.11, the DC force changed in increments of 1.3 N and 0.7 N. Yet the amount of variation due to large changes in DC Force was on the level of the intra and inter-vowel variability even though this subject kept the DC Force within ±0.11 N throughout the experiment. Indeed, these data suggest that the variability that was observed between trials could, at least in part be attributed to changes in the DC force.

A more extreme case is displayed in Figure 3.11 which shows two NTFs that were estimated using DC Forces that differed by 4.3 N. We see that the two NTFs in this figure look very similar despite the difference in DC Force, further supporting the claim that the NTF isn't dependent on DC Force.

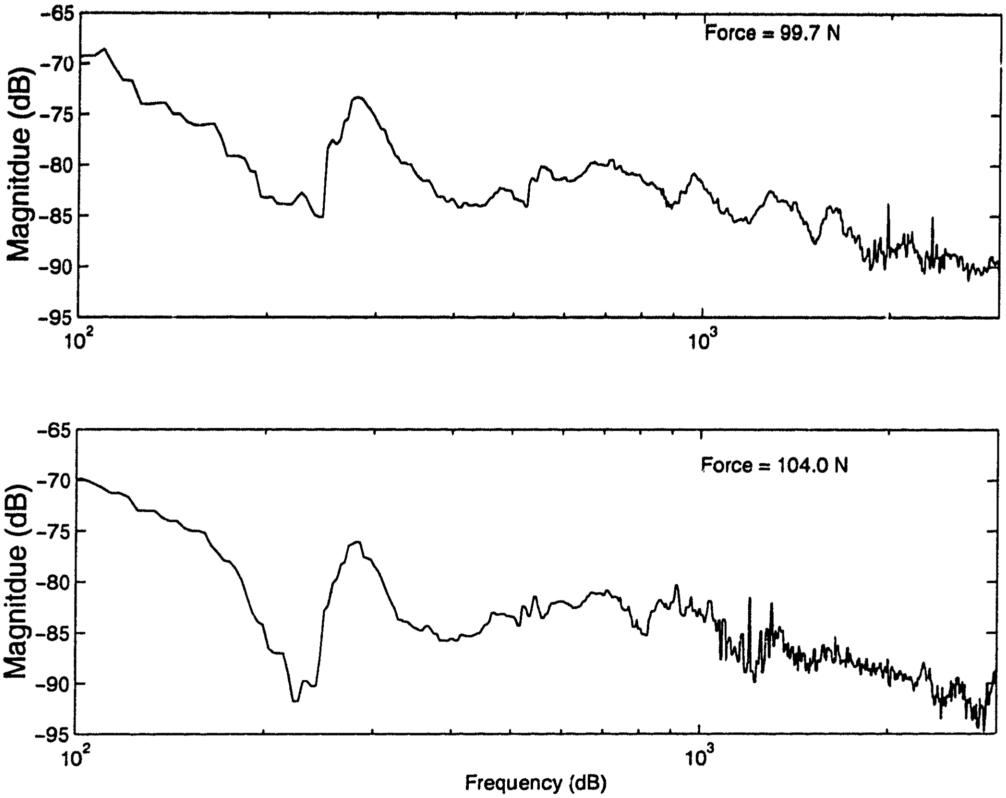


Figure 3.11. Two NTFs measured using DC Forces differing by 4.3 N. This was the largest force difference that was achieved (the differences in other subjects were between 0.6 and 2 N), yet the NTFs look strikingly similar.

Another aspect to consider is how steadily the subjects held the shaker in place during an experimental trial. It was discovered that over time, the most a subject changed the amount of force was by ±0.22 N, although it was more common to find a change on the order of ±0.1 N. These values closely resemble those associated with the trial to trial variability and are comparatively smaller than the changes in force incurred when the subjects were varying it on purpose. Given the small amount of change in the NTF that occurred when large force changes were applied to the neck, it is unlikely that the small variations in the DC force that occurred during a trial had an appreciable effect on the NTF.

IV. Discussion

A. Implications of the Results

The goal of this research was to measure the neck transfer function of both laryngectomy patients and people with their larynges intact. Specific goals were to understand how the NTF varied across subject groups (males vs. females, and laryngectomees vs. laryngeals), and over different locations on the neck wall. The results indicate that there was little difference between the mean NTF of all laryngectomized subjects and the mean NTF of all laryngeal subjects. However, when the laryngeal and laryngectomy groups were subdivided into male and female groups, it was found that while the mean NTFs of the male and female laryngectomy subjects were quite similar, the mean NTFs of the male and female laryngeal subjects proved to be different in several ways. The subjects were also grouped so that comparisons were made between laryngeal and laryngectomized subjects of the same sex. These comparisons suggested that although the laryngectomized and laryngeal males had comparable mean NTFs, the laryngectomized and laryngeal females had notably distinct NTFs.

The lack of significant differences between the male laryngectomy and laryngeal NTFs is somewhat puzzling, especially given the differences present in the female laryngeal and laryngectomized NTFs. Every laryngectomized subject had undergone a total laryngectomy and thus none of them had any residual laryngeal tissue in the neck. Furthermore, all of them had undergone radiation therapy in addition to the laryngectomy and thus it was expected that the radiation would change the properties of the skin to some degree. In sum, it was thought that the neck tissue characteristics of laryngectomized subjects would be different from those of the normal subjects and, accordingly, it was expected to find notable differences in the NTFs. As was discussed in Section II-D, the set of signals that were collected for each experimental trial had to meet a set of minimum quality criteria. Any set of signals that did not meet these criteria was not analyzed, thereby effectively removing any neck tissue with poor transmission qualities from the analysis. The only data that were discarded were collected from laryngectomized subjects presumably because the neck tissue where the recordings were made did pass enough sound into the vocal tract such that a good recording could be made. Since none of the male laryngeal subjects produced data that was discarded, it could be hypothesized that because of the quality constraint on the data, the only laryngectomee data that was analyzed was recorded at neck locations that had transmission properties that were similar to those of the necks of the normal subjects. Hence, because the transmission properties of the necks of both subject groups that were analyzed were similar, it should not be unexpected to find that the transfer functions of the laryngeal and laryngectomized subjects are similar.

Yet, this explanation does not account for the fact the female laryngeal NTFs differ from those of all other subject groups. It is possible that anatomical-morphological differences produced the different NTFs. However, only 2 laryngeal female subjects were tested and therefore the conclusions that can be drawn from these differences are limited. Indeed it is possible that the NTFs of the two subjects that were measured did not represent the average characteristics of the normal female population, i.e. sampling error.

While the differences found for the normal female NTF are interesting, it is important to keep in mind that the driving motivation for this research is the improvement of the current state of electrolarynx technology. NTF measurements were made so that these estimates could be used to design an optimal electrolarynx driving signal. Therefore the NTFs of the laryngectomized group are of primary importance. Based on the data in Figure 3-10b and Table 3-5, the only significant difference in the NTFs of male and female laryngectomees is in the low frequency slope (-12.6 dB/octave for females vs. –8.3 dB/octave for males), and therefore an EL driving signal designed for a female user may need more of a low frequency boost than a driving signal for a male user.

Another point that must be considered when designing a new driving signal is the inter-subject variability of the NTF. In other words, are the average NTFs which were computed for the laryngectomized subjects useful for designing a new EL driving signal for all laryngectomy patients, or should NTF measurements be performed on each individual EL user and the driving signal designed according to that particular user's NTF. Although many of the transfer functions resemble those displayed in Figure 3.10b, some subjects had NTFs that were noticeably different. To illustrate the individual variability, Figure 4-1 displays the average NTFs at the position with the greatest overall gain for three laryngectomized males.

Subject *ssw*'s NTF actually looks to have a flat response, unlike the low pass average laryngectomee NTF. Thus it appears likely that using the average male laryngectomee NTF to design an EL driving signal for this subject would not be prudent. In the case of subject *jfs,* the NTF has a low frequency slope that is over -5 dB/octave steeper than the average. If the average NTF were to be used to construct a driving signal for this subject, there might be a deficit in the sound energy between 100 and 300 Hz in the resulting electrolaryngeal speech. The NTF of subject *rsm*, on the other hand, closely resembles the average NTF, and of these

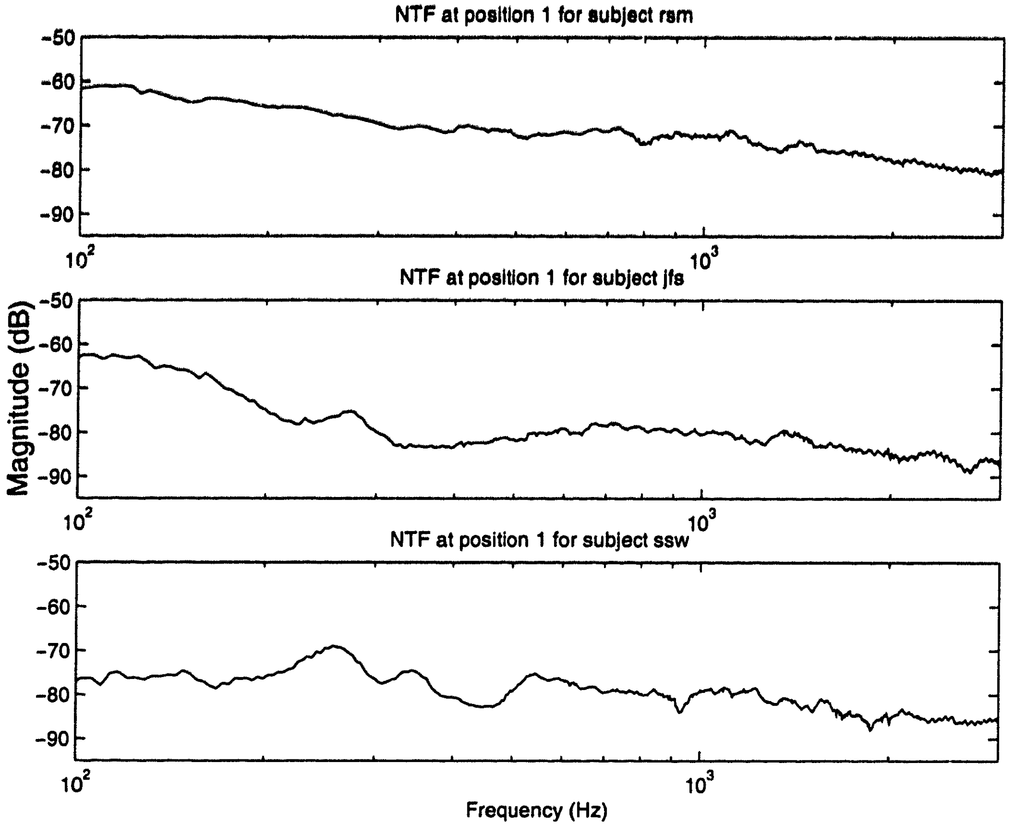


Figure 4.1. The NTFs of three laryngectomized males show that there can be a noticeable difference between the NTFs of individual subjects, despite those subjects being members of the same subject group. The NTF of subject rsm looks very similar to the average laryngectomee NTF (in Figure 3.7), while the NTF of subject jfs has a steeper low frequency slope (-15 dB/octave) than the average (-9.9 dB/octave). In addition, the NTF of subject jfs has a positive slope between 400 and 700 Hz (rather than a flat slope). Subject ssw's NTF is extremely different from the average NTF (and the NTFs of the other subjects), and looks to have a fairly flat response.

three subjects, he would gain the most benefit of using the average NTF to design the EL driving signal.

The large differences in these three subjects' NTFs suggest that while some general descriptions about the NTF can be made, there is a certain degree of inter-subject variability. While it may be beneficial to use the average NTF to design an electrolarynx for some electrolarynx users (e.g. subject rsm), in others, it would be useful to make an NTF measurement of the EL user. With such a measurement, an EL driving signal can be designed to compensate for the features present in an individual's NTF. For example, subject *jfs* would need the electrolarynx to provide extra energy to compensate for the steep negative slope below 300 Hz, while subject *ssw* would need frequencies below 250 Hz to be enhanced.

The effect on the NTF of changes in DC Force is also important when using an NTF measurement to design an electrolarynx. Each user holds the electrolarynx against his/her necks with different applied pressures, and the amount of pressure used by an individual can vary over time. Therefore, if the NTF were dependent on the pressure (or DC Force, since pressure is simply Force/Area), trying to design an EL driving signal using the NTF would be vastly more complicated. Fortunately, the results demonstrate that the NTF is not significantly affected by changes in the applied pressure on the neck.

The lack of DC Force dependence of NTF is promising for attempting to improve the electrolarynx in another fashion. In addition to its artificial quality, one of the most blatant shortcomings of EL speech is its lack of pitch variation. Thus any semblance of pitch control that can be added to an electrolarynx would greatly improve the perceived quality of EL speech. In one attempt to add pitch control, some moderate success was achieved by placing a pressure sensor on the cap of an electrolarynx so that the user can adjust the pitch by changing the applied pressure to the neck (Goldstein 1998). Because the NTF is independent of the pressure applied to the neck, there is no reason to be concerned about the EL driving signal needing real-time adjustment (to compensate for a time varying NTF) as the user varies the pressure to change the pitch.

It has been well documented (Weiss *et al.*, Qi and Weinberg) that electrolaryngeal speech contains a low frequency deficit, i.e. at frequencies below 500 Hz, there is significantly less energy in EL speech than there is in normal speech. Figure 4-2 shows an example of the low frequency deficit in the vowel /ae/. Tests done on a Servox electrolarynx demonstrate that a great deal of the low frequency deficit stems from the excitation that the Servox provides (Houston, unpublished). However it was important to determine if the neck tissue also contributes to the deficit.

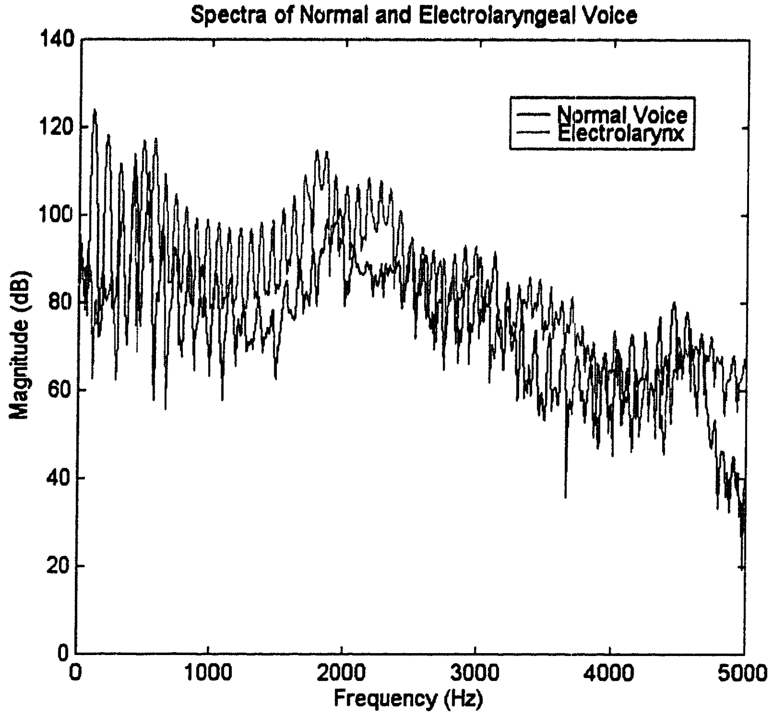


Figure 4.2. Plots of the spectra of the vowel /ae/ spoken by a laryngeal male (subject gsm) with his normal voice and with a Servox electrolarynx. The microphone was located 2 cm from the lips. The low frequency deficit is apparent below 500 Hz The energy at 100 Hz is about 35 dB less than the energy at 500 Hz for the EL speech, while for normal speech, the energy at around 100 Hz is actually 10 dB greater than the energy at 500 Hz.

From the results of this research, it obvious that the neck does not contribute to the low frequency deficit but in fact helps (to a some degree) to equalize the signal (produced with a currently available EL, e.g. the Servox) which ultimately excites the vocal tract.

B. Comparison with other research

One of the stated goals of this research was to extend the work done by Norton and Bernstein. Although the main results of this thesis concern the NTFs of the laryngectomees the NTFs of the laryngeal group (especially the males) can be used to compare with the results of their study. Norton and Bernstein followed a similar experimental protocol although they did not make closed mouth measurements in an attempt to remove the directly radiated noise, and they did not estimate and remove the lip radiation characteristic. Furthermore, their transfer function (or FRF as they have named it) is the ratio of the estimated volume velocity to the input voltage to the shaker thus making the FRF dependent on the shaker response. However, Norton and Bernstein used the same model B&K shaker that was used in this study, and as Figure 2.5 demonstrates, the shaker response is relatively flat (a change of about 6 dB over a freqency range of 3000 Hz). Thus the inclusion of the shaker response in the FRF cannot be an explanation as to why the Norton and Bernstein FRF looks radically different from the NTFs of normal males reported in this thesis. According to Norton and Bernstein the FRF attenuates frequencies below 500 Hz and contains 4 sharp resonances between 600 and 3000 Hz. However when they used this FRF to derive a new EL driving signal (by using it to inverse filter a natural glottal wave form), the speech produced with the new driving signal was judged to be more pleasing than typical EL speech. The fact that they were able to use a transfer function that significantly differs from the ones reported here to produce an improvement in EL speech suggests that the NTFs reported in this thesis are incorrect. But if one considers the impact of Norton and Bernstein's FRF if it were inverted (i.e. to inverse filter the glottal source), it would boost the low frequencies and attenuate the higher ones, especially near the resonance peaks in the FRF (which become zeros in the inverse FRF). Thus the new EL driving signal would contain the necessary low frequency energy to make up the gap between EL and normal speech. However, as mentioned earlier, the Servox itself is responsible for the low-frequency deficit in EL speech, and hence, as Qi and Weinberg's (1991) work implies, if the Servox could produce the low frequencies present in normal speech, EL speech would sound more natural.

If the electrolarynx itself is responsible for the low frequency energy deficit, then why does the Norton and Bernstein FRF suggest that the neck tissue is responsible? This is because the sharp resonances in the FRF occur at higher frequencies and thus boost the high frequencies relative to those below 500 Hz. The question arises as to what is the significance of these resonances. The first three resonances have center frequencies of 650 Hz, 1100 Hz, and 2450 Hz, which are also very close to the average formant frequencies of the vowel /a/ spoken by a male (Peterson and Barney 1952). In fact if one compares the FRF resonances with the formants of the experimental subject saying the vowel /a/, one would see that they are located at virtually the same frequencies. Norton and Bernstein also needed a method of removing the vocal tract transfer function from the pressure signal measured at the lips. They chose to use cepstral deconvolution instead of linear prediction and inverse filtering as their method. Cepstral deconvolution is a useful technique for separating the vocal tract contribution to a speech signal from that of the excitation source. For voiced speech, this separation is relatively simple because the cepstrum contains a peak at the pitch period and by simply frequency-invariant high pass filtering the cepstrum, one can remove the vocal tract contribution from the signal. Unvoiced speech is more difficult because there is no peak at the pitch period. In the case of Norton and Bernstein's experiment, they drove the shaker with broadband noise, producing speech without a pitch period. It appears that Norton and Bernstein had the subject say the vowel /a/, found the time to perform the high pass frequency invariant filtering and then applied this filtering to the speech segment obtained by driving the neck with the shaker. The problem with this technique is two-fold. First, by using the same high pass filter as was used in the normal speech, it is quite possible that Norton and Bernstein filtered out some components of the neck transfer function. The cepstral deconvolution separated the part of the spectrum that is slowly varying in frequency from the rapidly varying part. Since it was expected that the neck transfer function would vary at least as slow as the vocal tract transfer function, and if the same high pass filter is used, it seems likely that the neck transfer function could have been filtered out with the vocal tract transfer function. In the research presented in this thesis, great care was taken to ensure that only the vocal tract resonances were removed from the pressure signal.

Second, as House and Stevens (1958) discovered, the bandwidths of the vocal tract formants are significantly narrower when the glottis is closed than when the glottis is open. The speech signal used to estimate the FRF was produced by having the subject hold his breath by closing his glottis) yet the frequency invariant high pass filtering location was determined using an open glottis speech signal. Therefore, the slow frequency varying part of the spectrum that is being removed has larger bandwidths than the actual slow varying part of the speech spectrum of the shaker driven speech segment. Thus the vocal tract resonances are not properly deconvolved, leaving the sharp peaks in the remaining signal spectrum. The end result is that it appears Norton and Bernstein's neck FRF differs from the NTFs reported here because they did not properly remove the vocal tract transfer function from the signal recorded at the lips.

Fujimura and Linqvist used a similar experimental protocol to measure vocal-tract characteristics, but instead of trying to remove the vocal tract transfer function from their measurements, they instead attempted to remove the neck transfer function. They derived a "correction function" to account for the transmission properties of the neck tissue which is similar to the NTF reported in this thesis. The correction functions computed by Fujimura and Linqvist also demonstrated a great deal of inter-subject and intra-subject variability, however, some of the correction functions do show some of the same features as the NTF's presented here. These correction functions show a peak at low frequencies, a decreasing magnitude until about

500 Hz where it flattens out and maintains a relatively flat spectrum until 1000 Hz where it starts decreasing up to 2000 Hz. In some cases a small peak could be seen between 300 and 500 Hz as well. However, the peak frequencies in the correction functions were found between 150 and 200 Hz rather than near the100 Hz location observed in the NTF's reported here. Furthermore, some of the correction functions look more high pass in nature, similar to the NTF of subject *ssw*. Nevertheless, the qualitative agreement between the correction functions of Fujimura and Lindqvist and the NTFs is encouraging.

It is also instructive to compare the transmission characteristics of the neck with transmission characteristics of other body tissues. Wodicka and Shannon (1990) measured the transfer function of the subglottal human respiratory system using 8 male subjects. They found that this transfer function is also low-pass, with a peak near 100 Hz and a minimum at 600 Hz (which was the limit of the frequency range they were able to measure). Wodicka and Shannon found resonances at 129 Hz and 143 Hz (depending on the measurement site) in the average transfer function for all the subjects, although these resonances were shown to be resonances in the chest cavity rather than a resonance from the tissue itself. Furthermore, the roll off of the subglottal transfer function was found to be -10±4 dB/octave and -17±5 dB/octave (between 300-600 Hz) (depending on the site on the chest), while the average roll off of the NTFs of the laryngeal subjects was measured to be –9.9±5.4 dB/octave (between 100-200 Hz). Although these slopes, (for at least at one position on the chest wall) are similar, the roll off for the chest wall transfer function continues up to higher frequencies than in the NTF (where the slope becomes flatter). It is not surprising to find a larger degree of high frequency attenuation in the chest transfer function due to the greater amount of tissue present in the chest, which would attenuate higher frequencies to a greater degree than would the neck tissue. Nevertheless, it is promising to find that the NTF qualitatively agrees with the subglottal respiratory system transfer function (in that they both act as low pass filters).

C. Preliminary modeling work – linear filter design

The primary use of the NTF is to aid in the design of an improved EL driving signal. While it is possible to use the empirical data presented in this thesis to develop the driving signal, it would be far easier to use an analytic transfer function in this process. To this end, some modeling has been done in the form of trying to match the poles and zeros of a linear filter such that the magnitude of the filter's frequency response looks similar to the measured NTF data. Although there are differences in the mean NTF of each subject sub-group, an inspection of them reveals that they all display a general trend. The NTFs have a low frequency roll off of about –10 dB/octave, then flatten out between 300 and 1000 Hz, and then roll off at about –5 dB/octave at frequencies greater than 1000 Hz. These frequency characteristics suggest that the NTF contains two low frequency poles near 100 Hz (which produce the low frequency roll-off), two zeros at a frequency near 200 Hz (to flatten out the NTF), and one high frequency pole at about 1000 Hz (which produces the high frequency roll-off). Therefore the goal was to find a transfer function of the form

$$N(s) = G\frac{(s-z_1)(s-z_2)}{(s-p_1)(s-p_2)(s-p_3)} \tag{14}$$

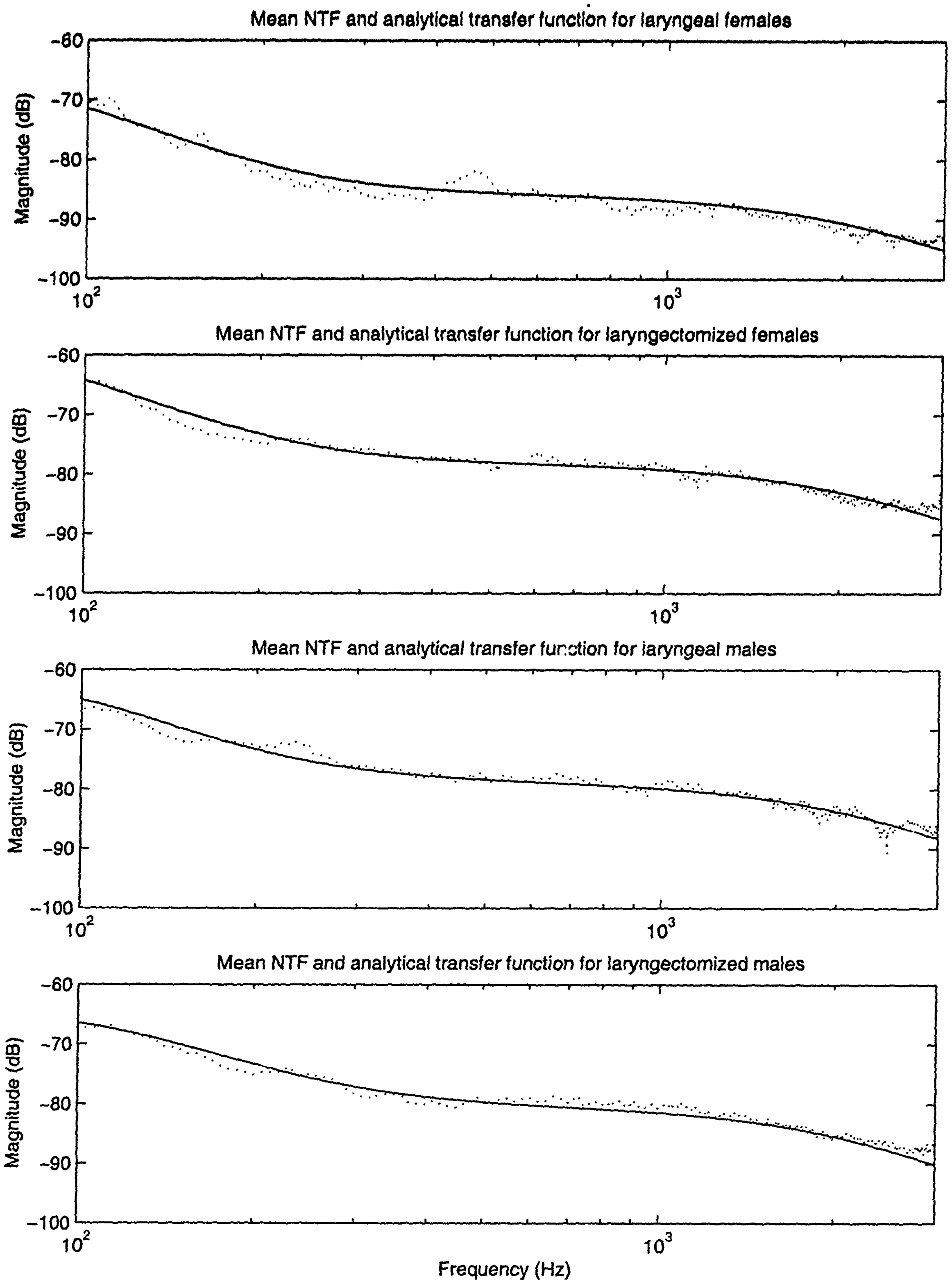
whose magnitude closely resembled that of the measured NTFs of the different subgroups, where G is the transfer function gain, the z_i are the zeros, and the p_i are the poles.

Starting with the baseline frequency locations for the zeros and poles mentioned above, the bandwidths and center frequencies of the poles and zeros, as well as the overall transfer function gain were adjusted such that the magnitude of $N(s)$ approximated that of the mean measured NTFs. While making the adjustments to $N(s)$, it was discovered that a fourth pole was needed to produce a good fit, and therefore *N*(*s*) has a complex-conjugate pair of zeros and two pairs of complex-conjugate poles. Four distinct *N*(*s*)'s were computed for the mean NTFs of the laryngeal and laryngectomized males and females and were compared with the empirical data. Figures 4.3 shows the mean NTFs for each subject group with the corresponding analytical transfer function while Table 4.1 presents the locations of the poles and zeros for each transfer function.

Feature	Laryngeal Females	Laryngectomized Females	Laryngeal Males	Laryngectomized Males
Gain	7,900	19,000	15,800	13,500
Pole Pair #1	$-100\pi \pm j120\pi$	$-100\pi \pm j100\pi$	$-100\pi \pm j160\pi$	$-300\pi \pm j160\pi$
Pole Pair #2	$-3400\pi \pm j2000\pi$	$-3400\pi \pm j2000\pi$	$-3400\pi \pm j2000\pi$	$-3200\pi \pm j2100\pi$
Zero Pair	$-350\pi \pm j300\pi$	$-350\pi \pm j300\pi$	$-350\pi \pm j240\pi$	$-900\pi \pm j240\pi$

Table 4.1. The poles, zeros, and gain for the analytical transfer function of each subject group. The gain is an arbitrary number used to make the magnitude of the model NTF correspond with the measured NTF. The complex values of the poles and zeros can be converted to frequency and bandwidth as follows: $f_c=|\mathrm{Im}\{p_i\}/2\pi|$ and $B_w=-\mathrm{Re}\{p_i\}/\pi$.

In all four cases, the analytical function provided an excellent fit, and for three out of the four cases, the maximum deviation of the model transfer function from the measured NTF was less that 3 dB. For the laryngeal males, there appears to be a spurious point at 2,450 Hz which produced a maximum deviation of 5.0 dB. If this point is excluded, the maximum difference is 2.4 dB. The data presented in Table 4.1 demonstrates that the original estimates (which were made based on the shape of NTF curve) of the locations for the poles and zeros were not too different from the final values, although the location of the zero pair is closer to the location of the first conjugate pole pair than expected (150 Hz for the females and 120 Hz for the males). Furthermore, based on the locations of the poles and zeros, it would appear that of the four transfer functions, the male laryngectomy one is the most distinct, especially in terms of the bandwidths of the poles and zeros. This contradicts the empirical finding that the female laryngeal mean NTF was the most dissimilar, (although it should be noted that the gain for model NTF for the laryngeal females was much smaller than those of the other NTFs as was found in the measured NTFs). However, caution must be taken when making conclusions from the model NTFs for two reasons. First, since only the magnitude of the frequency response was specified (and only over certain frequency range) there are an infinite number of valid transfer functions that can be used as a model. Second, the center frequencies and bandwidths of the poles and zeros were chosen based on how well the resulting transfer function appeared to match the measured NTFs, i.e. the decision criteria were subjective. A more objective set of criteria, such as minimizing the mean squared error of the magnitude of the frequency response, and possibly applying different weights to different parts of the frequency range would eliminate human subjectivity and strengthen any conclusions drawn from the modeling. This would be an obvious next step in the modeling process. Nevertheless, the current analytic transfer functions are still useful as they do appear to fit the measured NTFs quite well. From these functions, a digital filter can be designed (by means of the bilinear transformation or other method) and used to inverse filter a natural glottal sound source to produce a new electrolarynx driving signal. Another possible use is to design a circuit that has the same frequency response as $1/N(f)$ and incorporate this circuit into the electrolarynx itself. With such a circuit in place, the various components could be adjusted to compensate for the variability in the NTFs associated with each individual user. How the transfer function will be used in the electrolarynx design will ultimately depend on which method is simpler and produces a more natural sounding voice.


Mean NTF and analytical transfer function for laryngeal females

Mean NTF and analytical transfer function for laryngectomized females

Mean NTF and analytical transfer function for laryngeal males

Mean NTF and analytical transfer function for laryngectomized males

Figure 4.3. The mean NTFs for each subject group with the corresponding analytical transfer function fit. In all cases, the analytic function provided a good fit, usually producing a maximum deviation of less than 3 dB. All of the transfer functions consisted of two conjugate pole pairs and one conjugate zero pair

D. Limitations of this research

The experimental protocol that was employed in this study was chosen, in part, because of its non-invasive nature. However the non-invasiveness comes with a price. First, as was mentioned earlier, the only data that was analyzed from signals that contained a visible formant structure and had a coherence approaching unity. These criteria limited the locations on the necks of the laryngectomy patients at which reliable measurements could be made and thus were probably constrained to only measuring neck locations that had similar sound transmission properties. Many laryngectomy patients had locations on their necks that were not conducive for producing electrolarynx speech. It would have been of interest to measure the NTF at these locations to learn how the NTF of these locations differ (other than the apparent difference in overall gain) from the NTFs of those positions which transmit enough sound to produce audible EL speech. Measuring such potential different NTFs could provide some insight into how the NTF relates to the tissue present at certain locations on the neck. It is important to note, however, that with regard to improving the electrolarynx with these measurements, the only important location to measure is one that transmits a great deal of the sound into the vocal tract, as this will be the location at which the EL user will hold the device.

Figure 2.4 shows the pressure signal that is measured by the microphone at the lips is corrupted by the sound pressure that is generated by the shaker and transmitted through the air. The cylindrical silencer was created to help reduce the intensity of the directly radiated shaker sound, and while it was effective across some frequencies (especially at higher frequencies) at others, it was not as effective, as shown in Figure 2.3. As a result, for certain frequencies, the directly radiated sound was almost at the same level as the sound output from the subjects' mouths.

To illustrate this point, Figure 4.4 plots the magnitude of $\frac{S_{pv}(f)}{S_{vv}(f)}$ (or equivalently $T_1(f)\cdot N(f)\cdot VT(f)\cdot R(f)+T_2(f)\cdot H(f)$) and the magnitude of $\frac{S_{ev}(f)}{S_{vv}(f)}$ (or equivalently $T_2(f)\cdot H(f)$) computed from open and closed mouth measurements respectively. These data were obtained from a male laryngectomized subject. Ideally, $\frac{S_{ev}(f)}{S_{vv}(f)}$ should be smaller than $\frac{S_{pv}(f)}{S_{vv}(f)}$ because as part of the analysis algorithm $\frac{S_{ev}(f)}{S_{vv}(f)}$ is subtracted from $\frac{S_{pv}(f)}{S_{vv}(f)}$ and if these two quantities are of similar magnitudes, then the remaining signal to be analyzed will be small, thus limiting the accuracy of the estimated NTF. As Figure 4.4 shows, for certain frequencies, $\frac{S_{ev}(f)}{S_{vv}(f)}$ has a magnitude that is almost as large (or sometimes larger) than that of $\frac{S_{pv}(f)}{S_{vv}(f)}$ and at these frequencies, the NTF estimate is not as precise as it is at other frequencies.

Another drawback to using this experimental protocol is the need to estimate and remove the vocal tract transfer function from the signal measured at the lips. From an analysis point of view, measuring the sound pressure just on the other side of the neck wall (i.e. in the vocal tract) would be easier for estimating the NTF. Unfortunately, the required experiment would be highly invasive and difficult to perform. By using the experiment employed in this research, the difficulty arises in the estimation of the vocal tract transfer function. LPC was chosen to make the estimates because an all-pole filter is generally a good model of the vocal tract. In the case of this experiment (and in EL speech in general), this may not be the case.

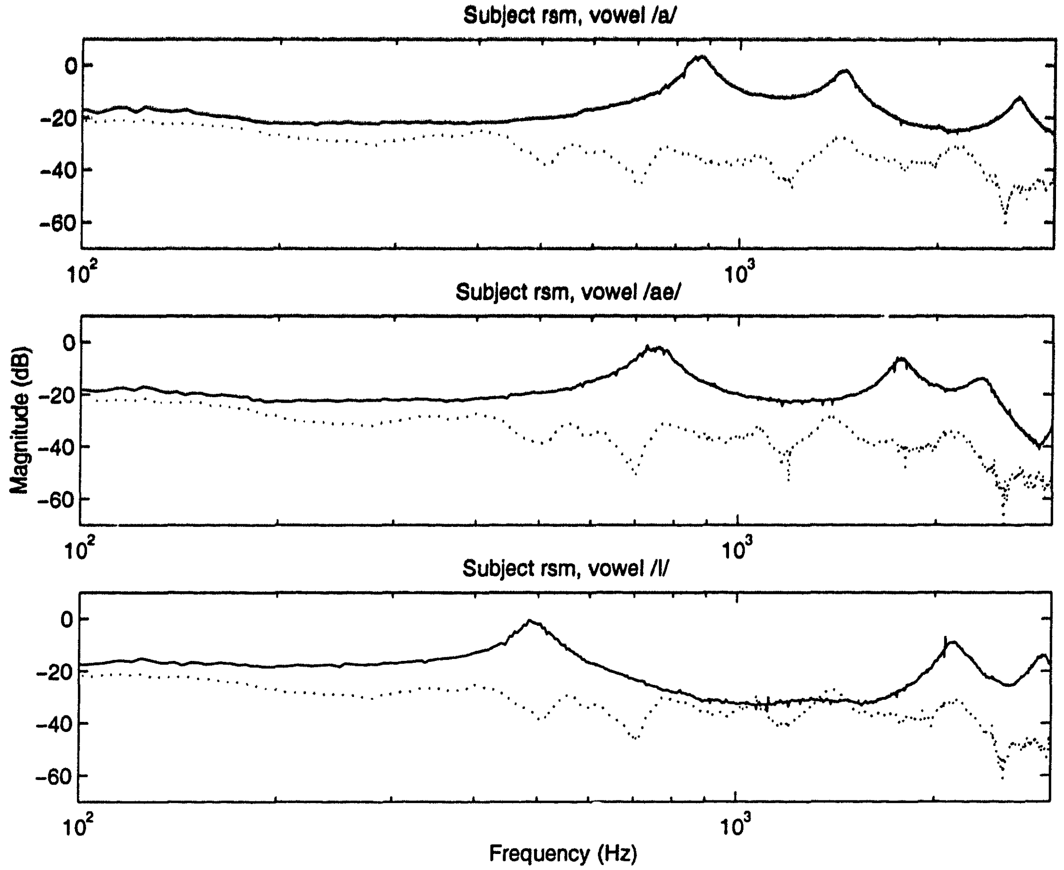


Figure 4.4. Plots of the magnitude of $S_{pv}(f)/S_{vv}(f)$ (solid lines) computed from open mouth measurements of 3 different vowels, along with plots of the magnitude of $S_{ev}(f)/S_{vv}(f)$ (dotted lines) computed from the corresponding closed mouth measurements. For certain frequencies, the closed mouth transfer function has a magnitude that is close to (or greater than) that of the open mouth transfer function. It is conceivable for the magnitude of the closed mouth transfer function to be greater than that of the open mouth transfer function because both transfer functions are complex quantities.

When the shaker (or EL) is placed against the neck, it is driving the vocal tract at a location other than its terminal end, thus producing a back cavity. This back cavity introduces a zero into the spectrum of the vocal tract transfer function whose frequency location depends on the cavity's length. Thus, the vocal tract transfer function is no longer a true all pole filter. The effects of this extra zero can be examined by using a uniform tube as a model. Figure 4.5 displays a tube being excited somewhere other than its terminal end.

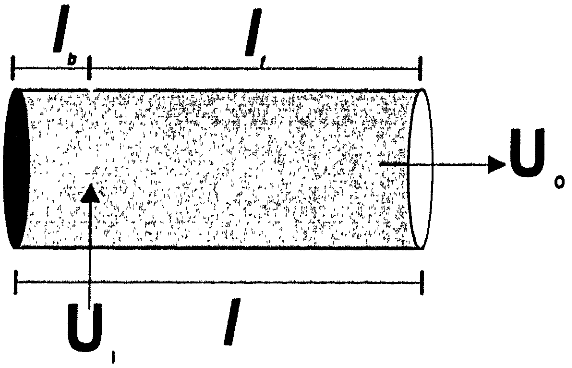


Figure 4.5. An open ended uniform tube being driven at a location in its middle. The length of the whole tube is denoted as *l*, the length of back cavity is l_b, and the length of the front cavity is l_f. U_I is the input volume velocity and U_O is the output volume velocity.

The transfer function of this tube (i.e. the ratio of the output volume velocity U_O to the input volume velocity, U_I) can be written as

$$\frac{U_O}{U_I}=\frac{-j\cdot\frac{\rho c}{A}\cdot\cot\left(\frac{2\pi fc}{l_b}\right)}{-j\cdot\frac{\rho c}{A}\cdot\cot\left(\frac{2\pi fc}{l_b}\right)+j\cdot\frac{\rho c}{A}\cdot\tan\left(\frac{2\pi fc}{l_f}\right)}, \tag{15}$$

where A is the area of the tube, ρ is the density of air, c is the speed of sound, l_b is the length of the back cavity and l_f is the length of the front tube. After some algebraic manipulation, this transfer function can be reduced to

$$\frac{U_I}{U_O}=\frac{\cos\left(\frac{2\pi fc}{l_b}\right)}{\cos\left(\frac{2\pi fc}{l}\right)}, \tag{16}$$

where l is the length of the entire tube. Therefore the first zero in this transfer function will occur where $\cos\left(\frac{2\pi fc}{l_b}\right)=\frac{\pi}{2}$, or where $f=\frac{c}{4l_b}$. For example, if the shaker is placed at a location that is 3 cm above the terminal end of the vocal tract, the zero will occur at about 2950 Hz. This zero will tilt the spectrum of the vocal tract transfer function down at higher frequencies: about -3 dB/octave between 1000 and 2000 Hz and –6 dB/octave between 2000 and 3000 Hz. Figure 4.4 plots the effects of the zero on the transfer function for a back cavity of length 3 cm.

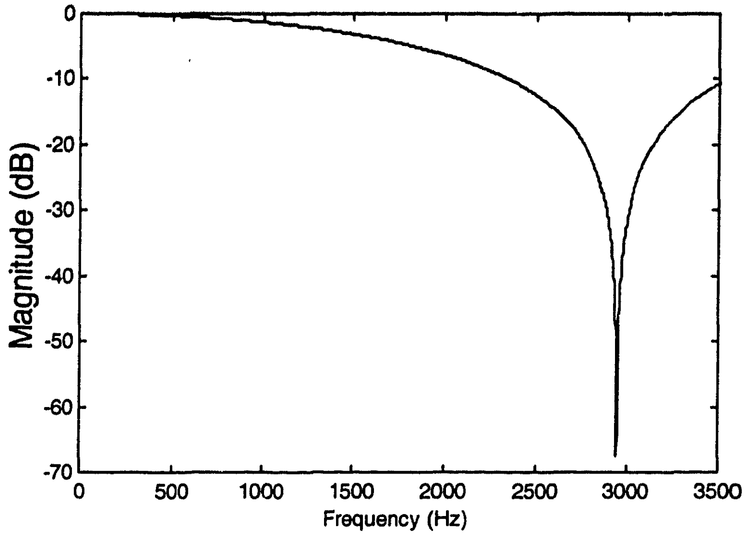


Figure 4.4. Plot of the effects of the zero produced by a back cavity of length = 3cm on the vocal tract transfer function. We see that at higher frequencies (near 2950 Hz), the effects are greater. It should be noted that this plot represents the behavior of an ideal model. In the actual vocal tract, the bandwidth of the zero would be greater.

At lower frequencies, there is little effect, but at higher ones, the spectrum is tilted downwards until it reaches the zero at 2950 Hz. This presence of the zero complicates the LPC fit, especially near the zero. Although LPC can often approximate the vocal tract transfer function when zeros are present (LPC analysis/synthesis is regularly used on nasal phonemes which contain zeros in their spectra), the presence of the zeros does complicate matters when correcting the LPC fit such that it only includes the vocal tract transfer function. However, when looking at the estimated NTFs, there do not appear to be any high frequency zeros in the estimated NTF, and it would seem that the LPC fit was able to account for the effects of the zeros. Finally, if the LPC fit were unable to accurately account for the zero and it affected the NTF, it may not be worth trying to correct the NTF for that zero. The ultimate aim of trying to improve EL speech is make it sound as close to normal speech as possible. Therefore if there is a zero in the EL spectrum, its effects must be removed in some fashion, so this task may be accomplished by keeping the zero in the NTF and designing an EL driving signal based on the NTF with the zero in it.

The statistical analysis performed on the data was intended to verify the trends that the spectra of the measured NTFs appeared to display. However, beyond this verification, the conclusions that can be drawn from the statistical analysis are limited by the small sample size. The subgroups that were compared consisted of as little as two subjects and therefore if any significant differences were found between subgroups (such as in the case for the normal female group) those differences may be due to the sampling error associated with a small sample size. If it were desired to further explore the differences that were found in the NTF then more subjects would have to be tested.

A model of the neck transfer function has not yet been developed. Such a model could provide some insight in to the NTF behavior, specifically the apparent sex dependence and lack of dependence of laryngeal status. This model could incorporate the physical properties of the neck tissue (e.g. mass and compliance) and relate these properties to the spectrum of the NTF. A first step has already been taken by matching a linear filter to the measured NTF data. Further work can use this filter to design an analogous circuit whose components (resistors, inductors and capacitors) can be related to the physical properties of the neck tissue. If this model could be related to the residual anatomy present in an EL user, an NTF measurement may not be needed for that user. At this point an obvious next step would be to develop a model of the NTF and use it to gain a greater understanding of what factors have the greatest influence in determining the NTF.

D. Summary

As a first step in improving the quality of electrolaryngeal speech measurements of the neck transfer function (NTF) have been made on nine laryngectomized and four laryngeal subjects. It was found that in general, the NTF peaks at or near 100 Hz, rolls off with a slope of about –10 dB/octave until 300 Hz where the magnitude of the slope decreases to zero slope until 1000 Hz, after which the NTF rolls off at about –6 dB/octave up to 3000 Hz. Many subjects' NTF displayed one or two small peaks between 200 and 500 Hz as well. When the subjects were separated into four groups (laryngeal females, laryngeal males, laryngectomized females, and laryngectomized males), it was found that the mean NTF of the laryngeal female significantly differed from the NTFs of the other three groups. Furthermore, despite the mean NTFs of the laryngectomized males and females being very similar, there was enough inter-subject variability to suggest that it would be useful (at least for some laryngectomees) to measure the NTF of an EL user so that the EL driving signal can be adjusted to compensate for the individual's NTF.

Finally, it has been found that the NTF can be modeled by an analytical transfer function comprised of two conjugate pairs of poles and one conjugate pair of zeros.

V. Appendix A

This section contains the average NTF for each position of each subject that was tested and subsequently analyzed.

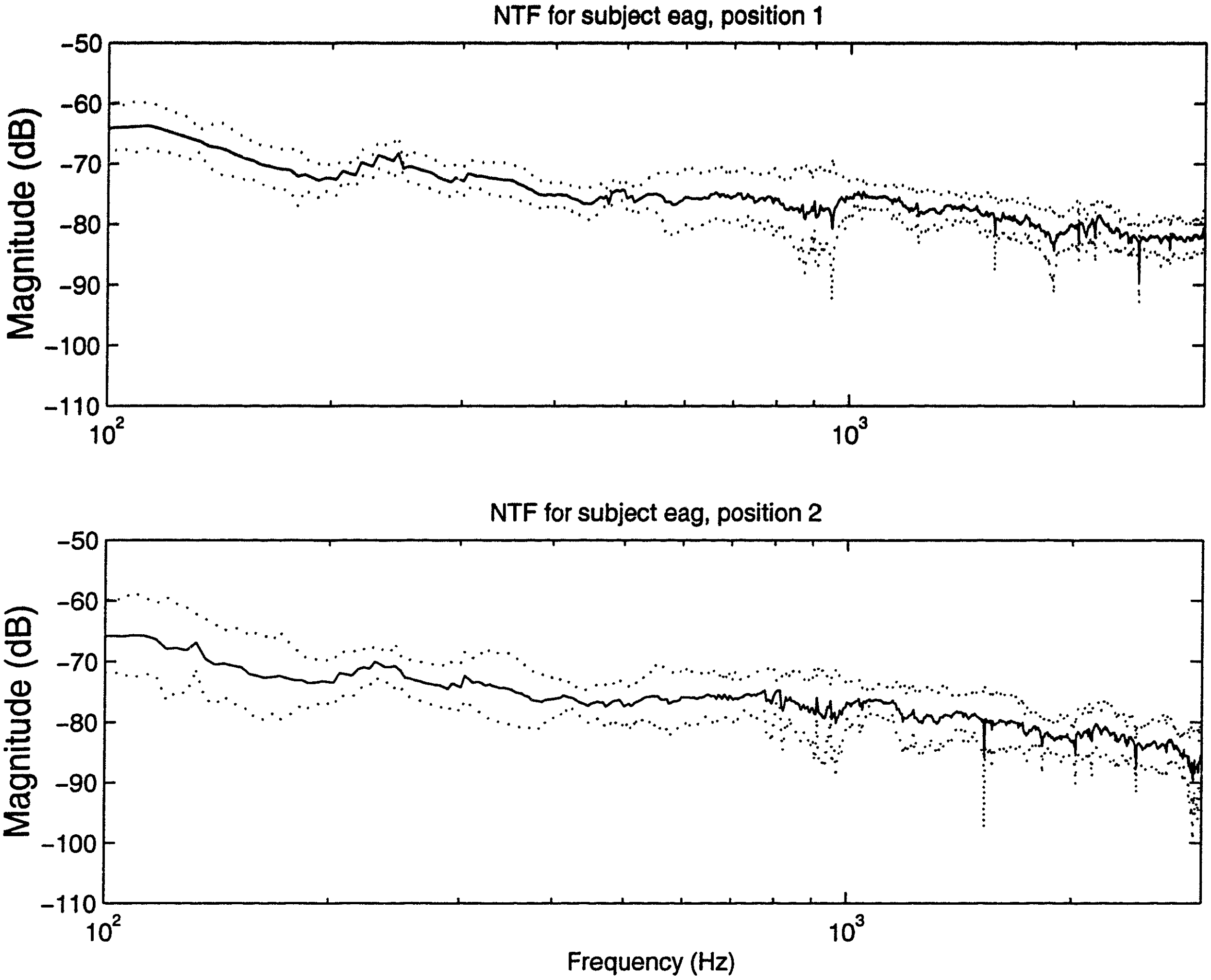


Figure A.1 The NTFs for subject eag, a male laryngeal subject. Only two positions were measured on this subject

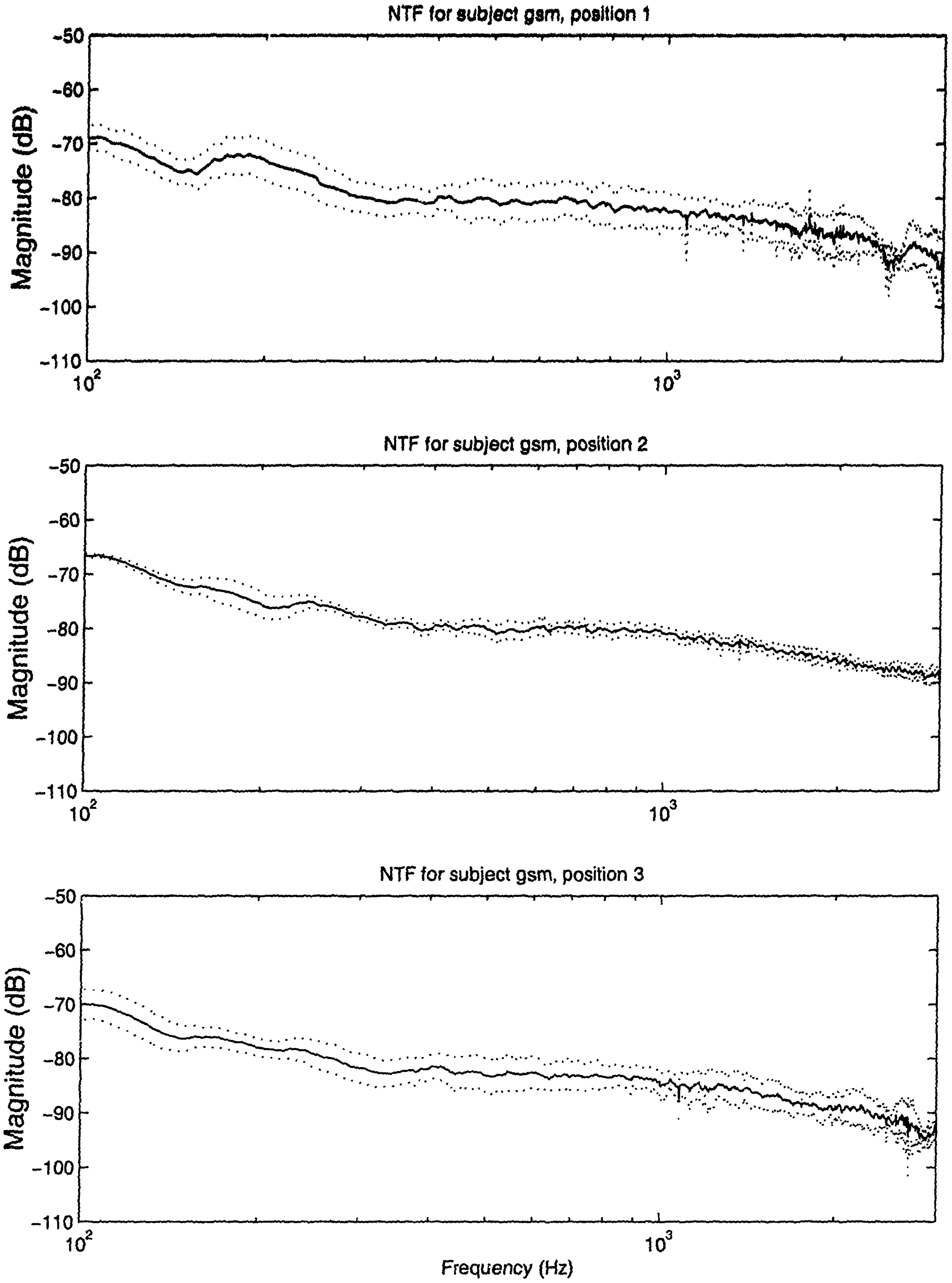


Figure A.2 The NTFs for subject gsm, a male laryngeal subject.

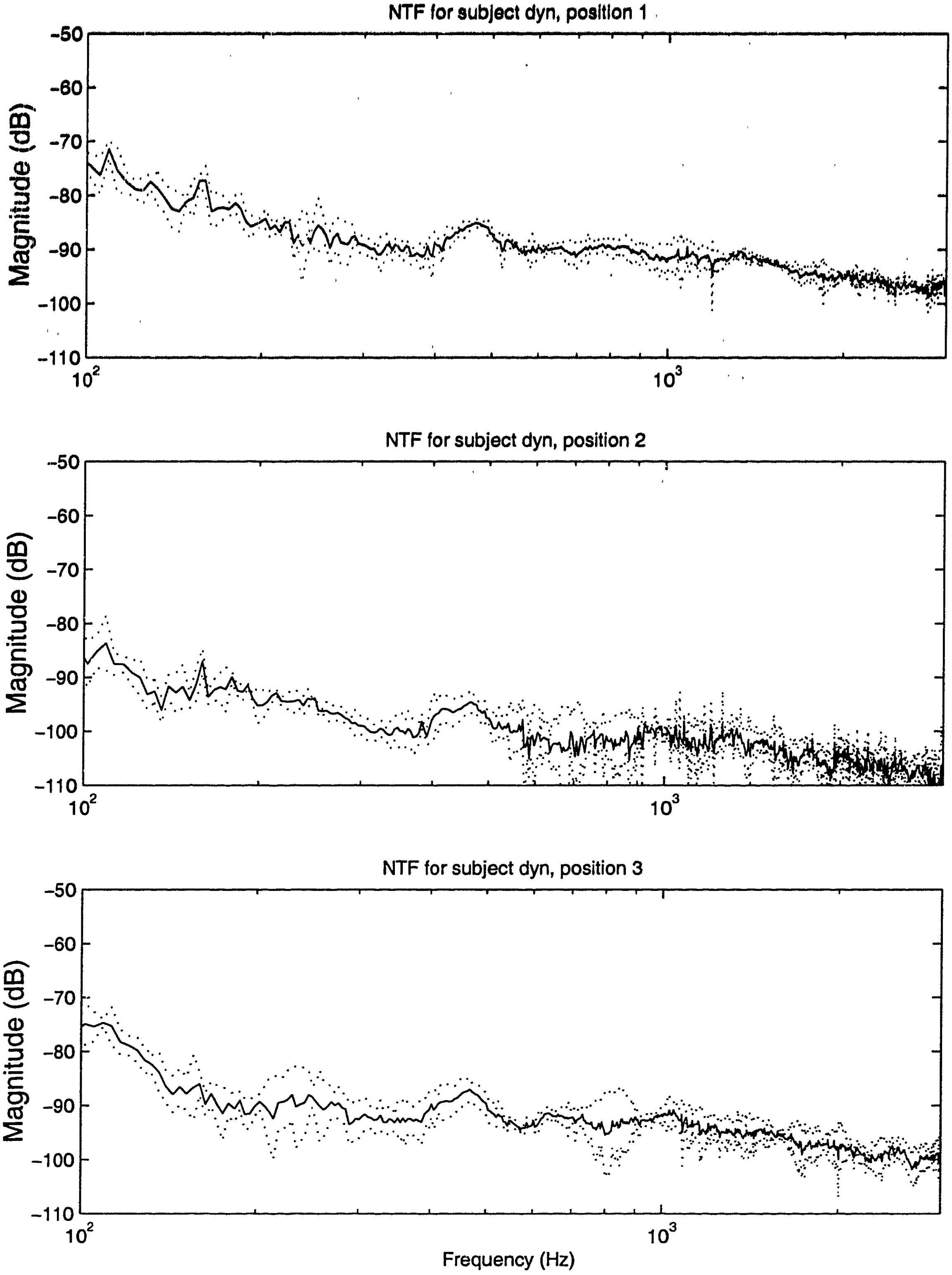


Figure A.3 The NTFs for subject dyn, a female laryngeal subject.

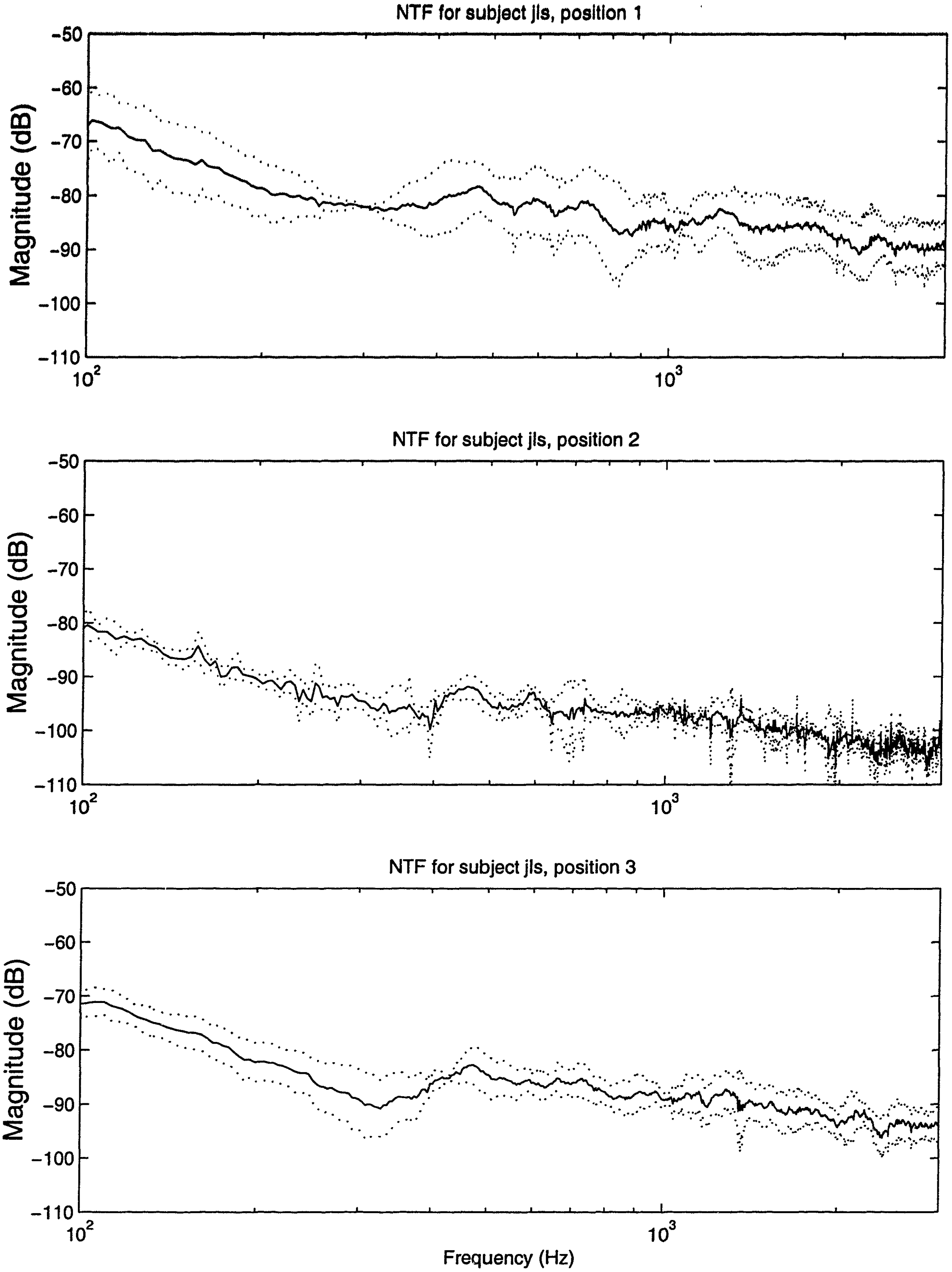


Figure A.4 The NTFs for subject jls, a female laryngeal subject.

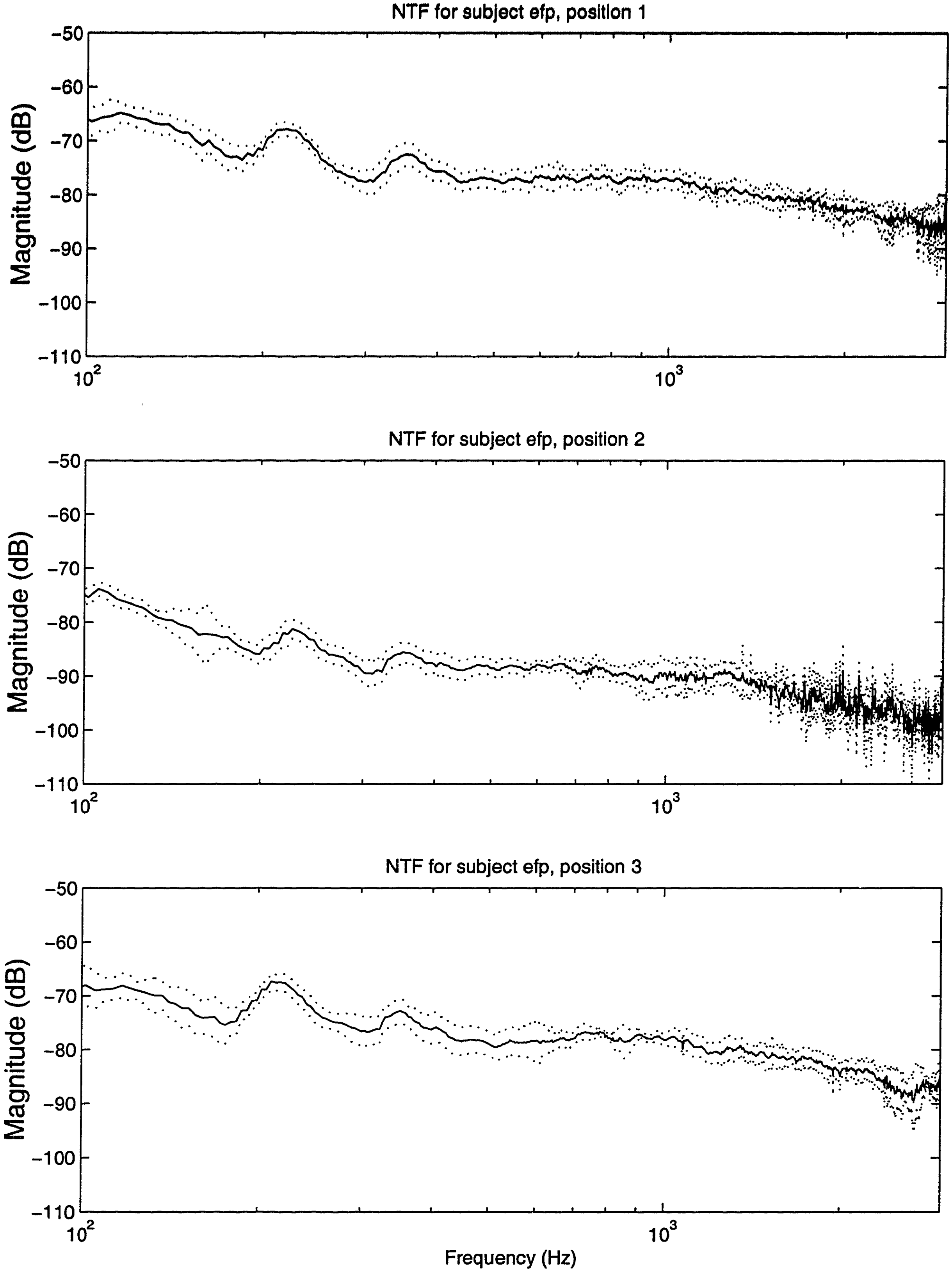


Figure A.5 The NTFs for subject efp, a male laryngectomized subject.

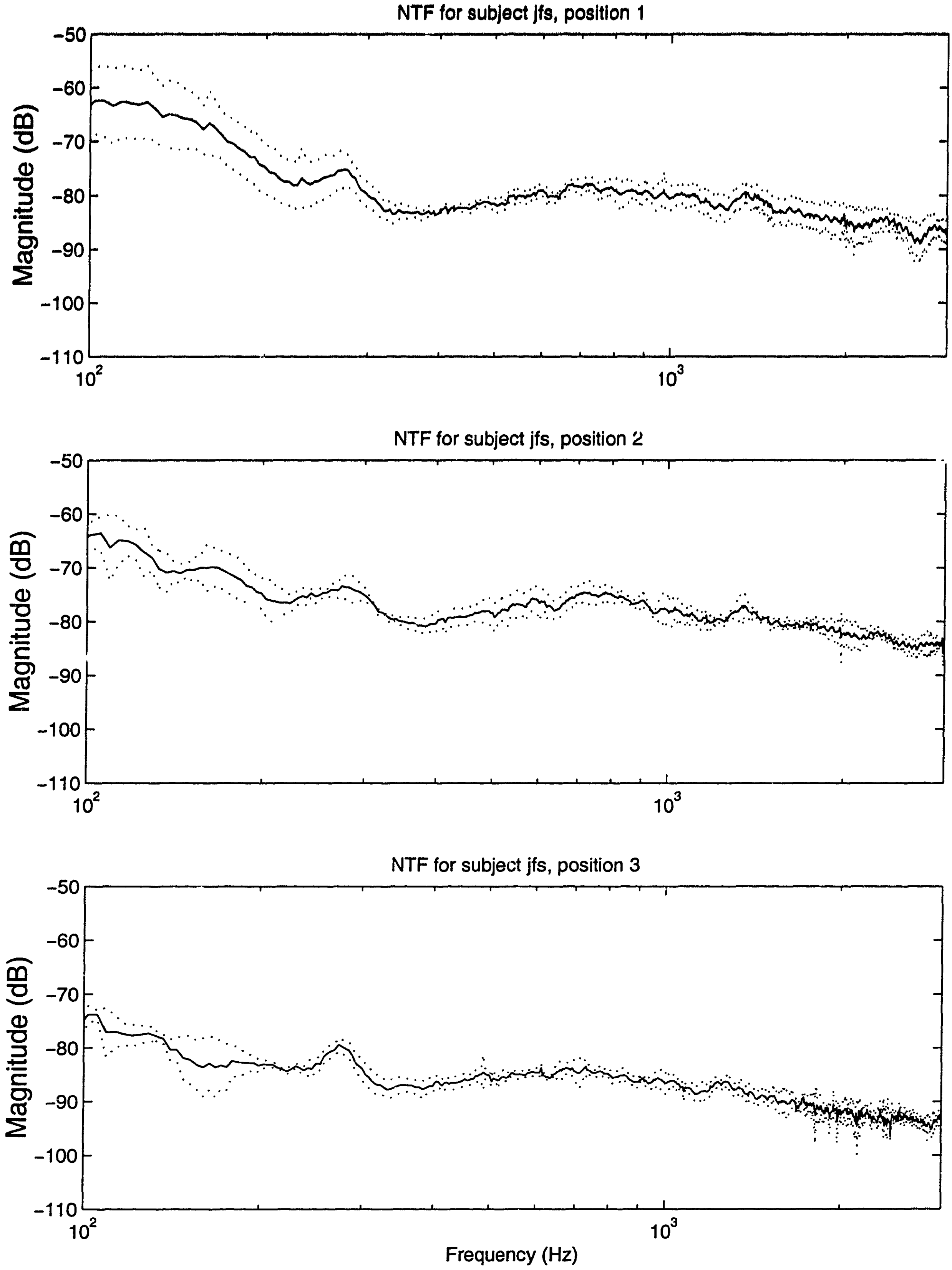
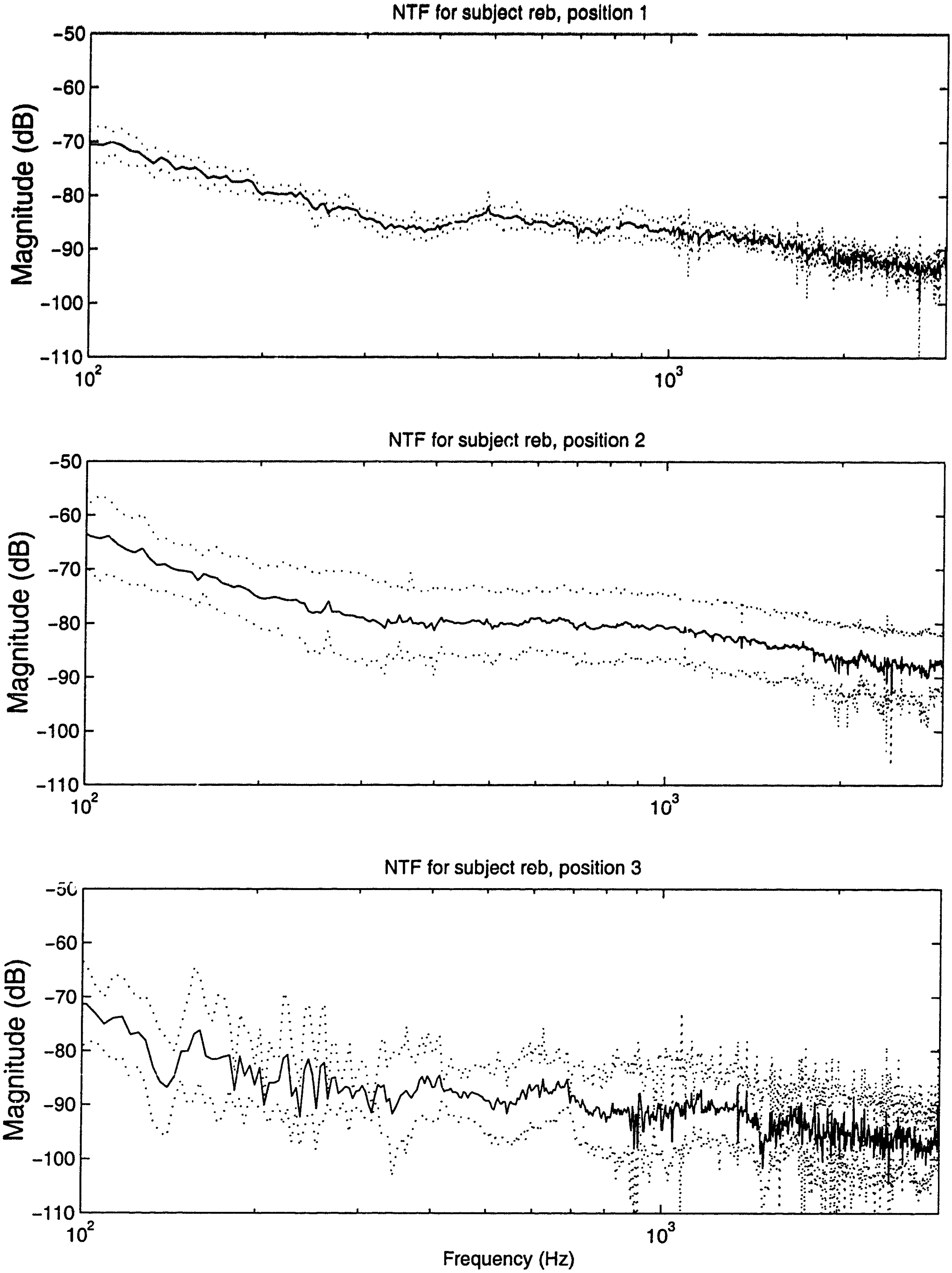


Figure A.6. The NTFs for subject jfs, a male laryngectomized subject.


NTF for subject reb, position 1

NTF for subject reb, position 2

NTF for subject reb, position 3

Figure A.7 The NTFs for subject reb, a male laryngectomized subject

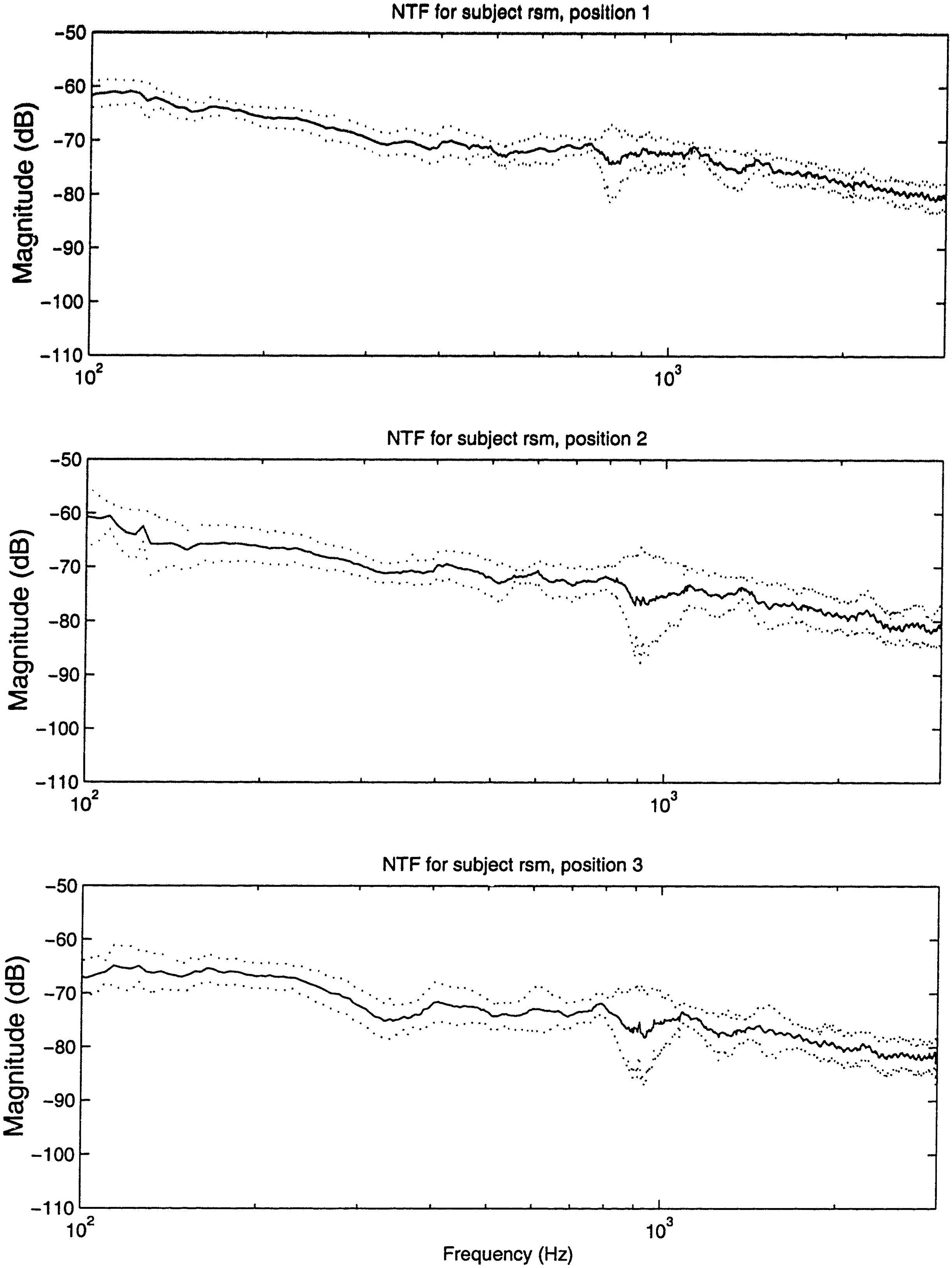


Figure A.8 The NTFs for subject rsm, a male laryngectomized subject.

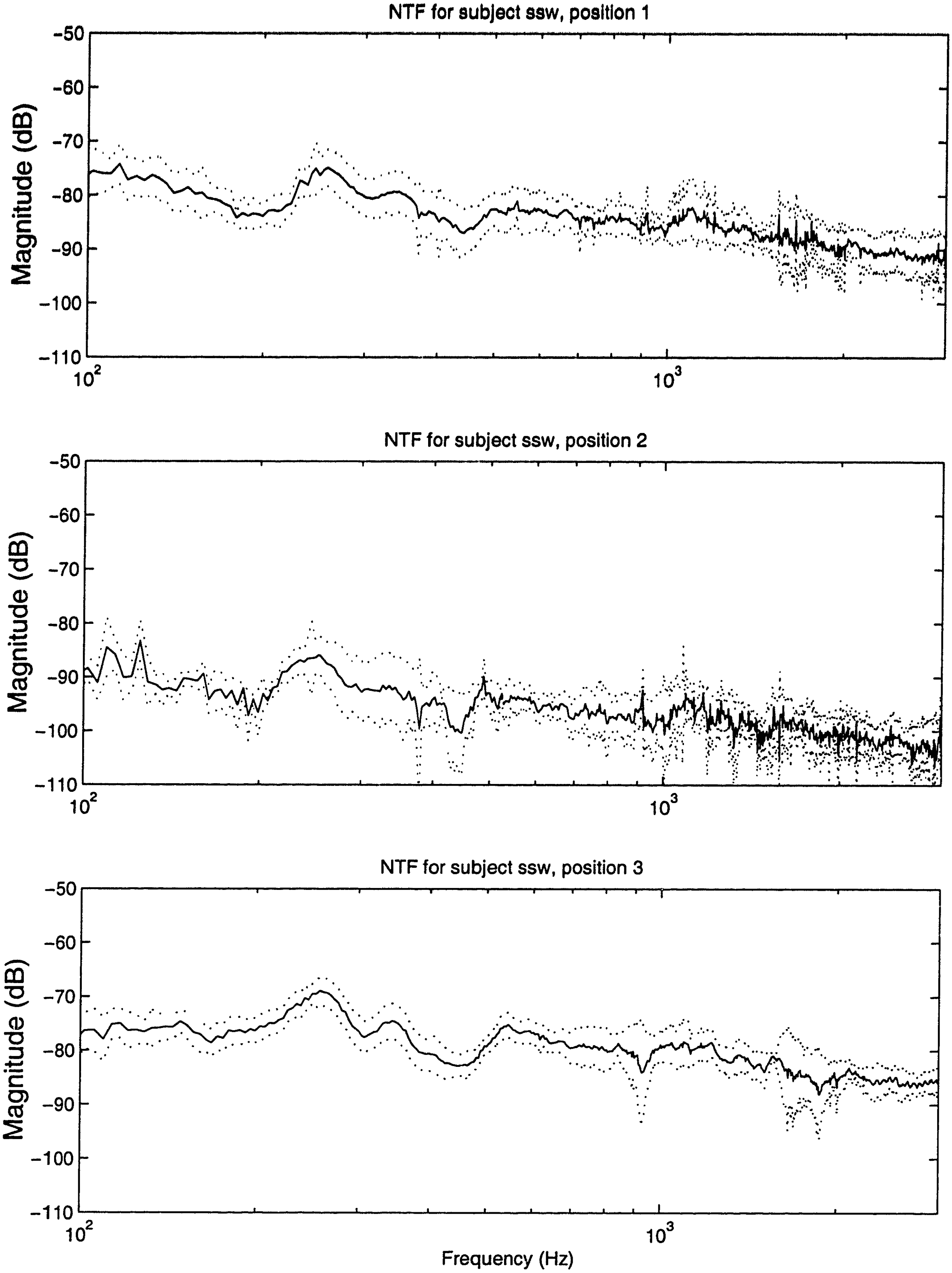


Figure A.9 The NTFs for subject ssw, a male laryngectomized subject.

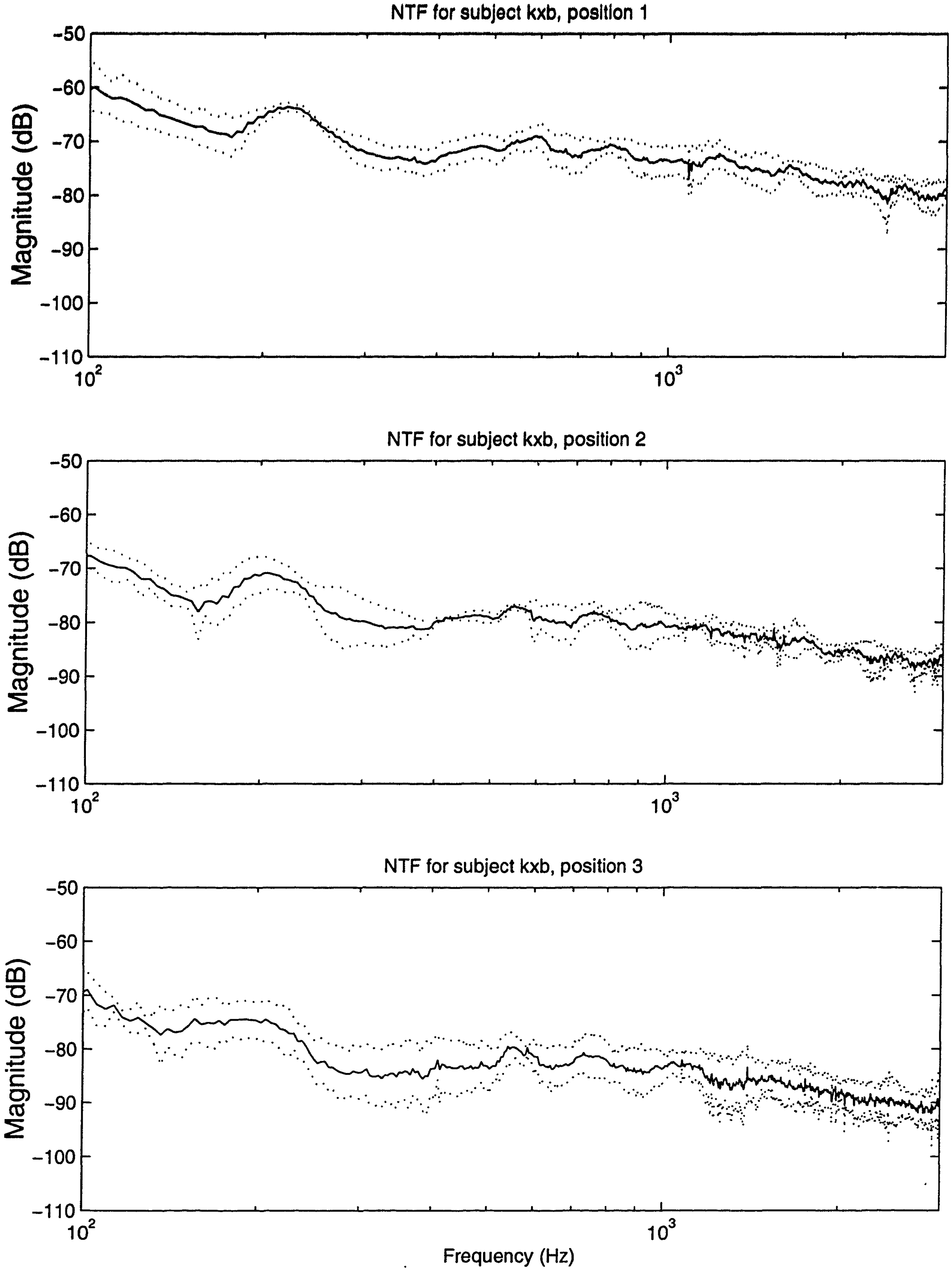
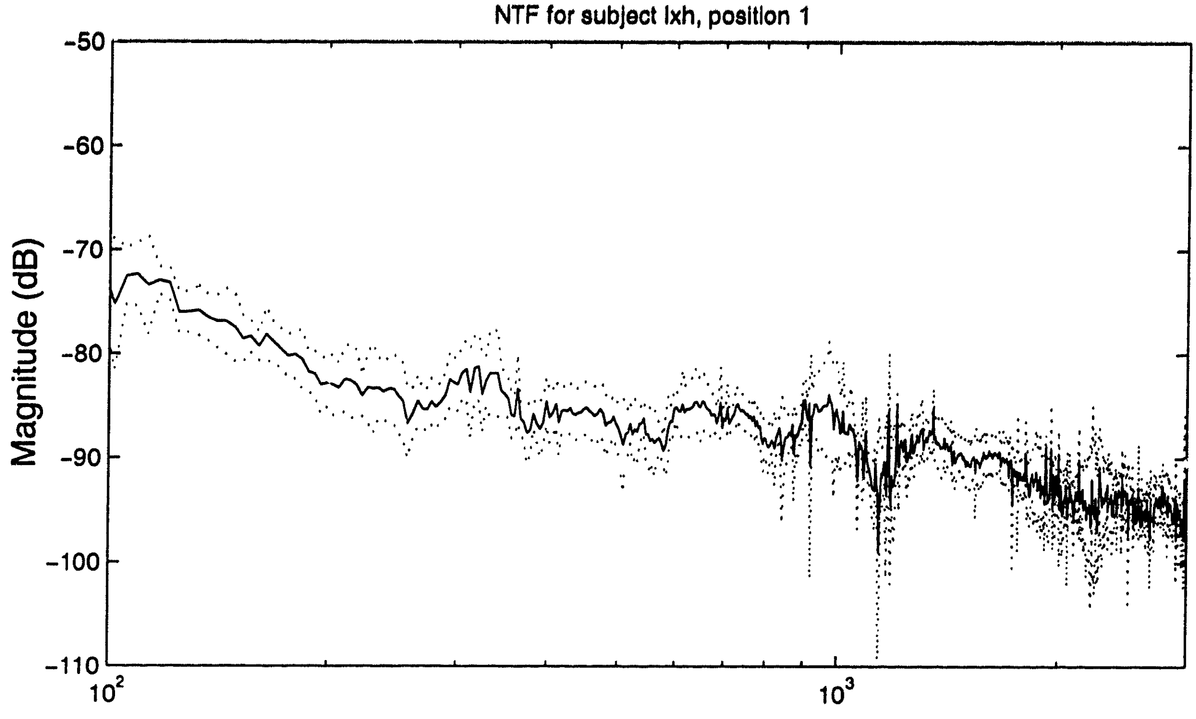


Figure A.10 The NTFs for subject kxb, a female laryngectomized subject.



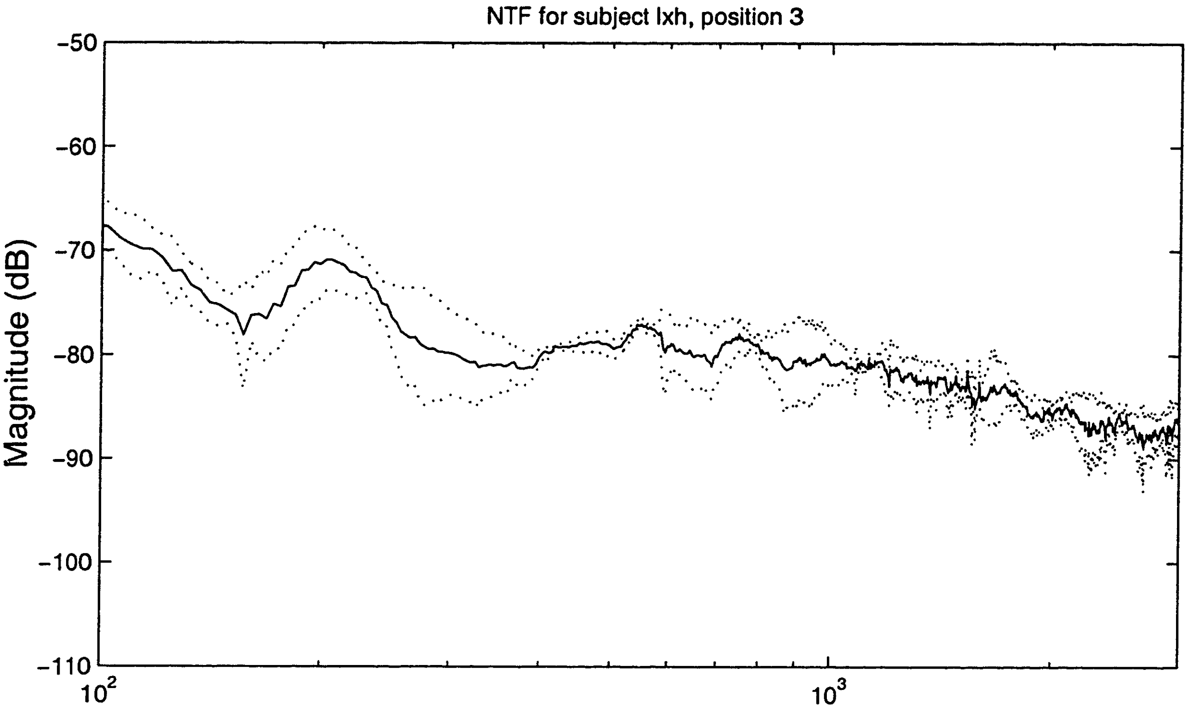


Figure A.11 The NTFs for subject lxh, a female laryngectomized subject. Measurements were only able to be made at two positions.

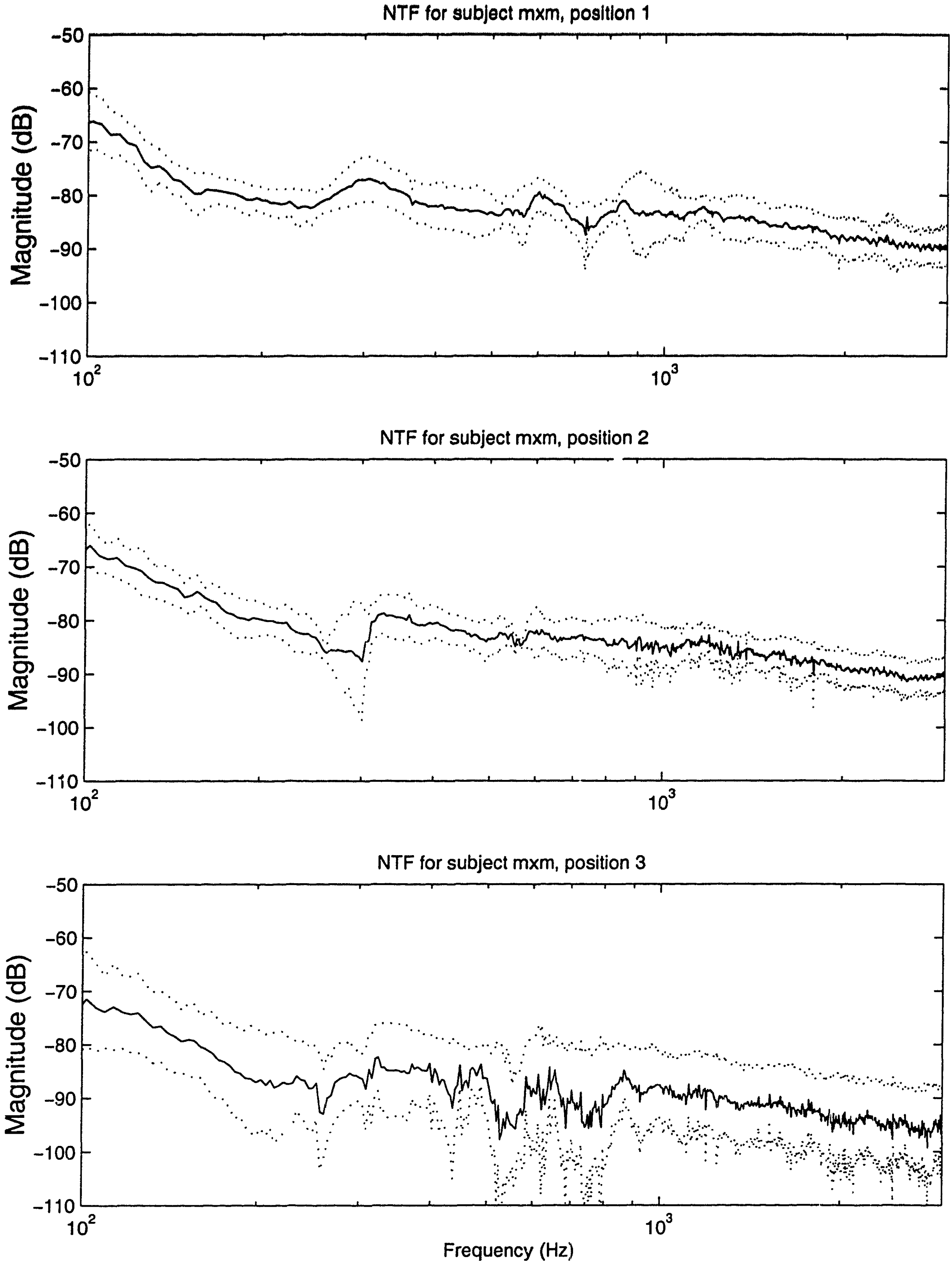


Figure A.12 The NTFs for subject mxm, a female laryngectomized subject.

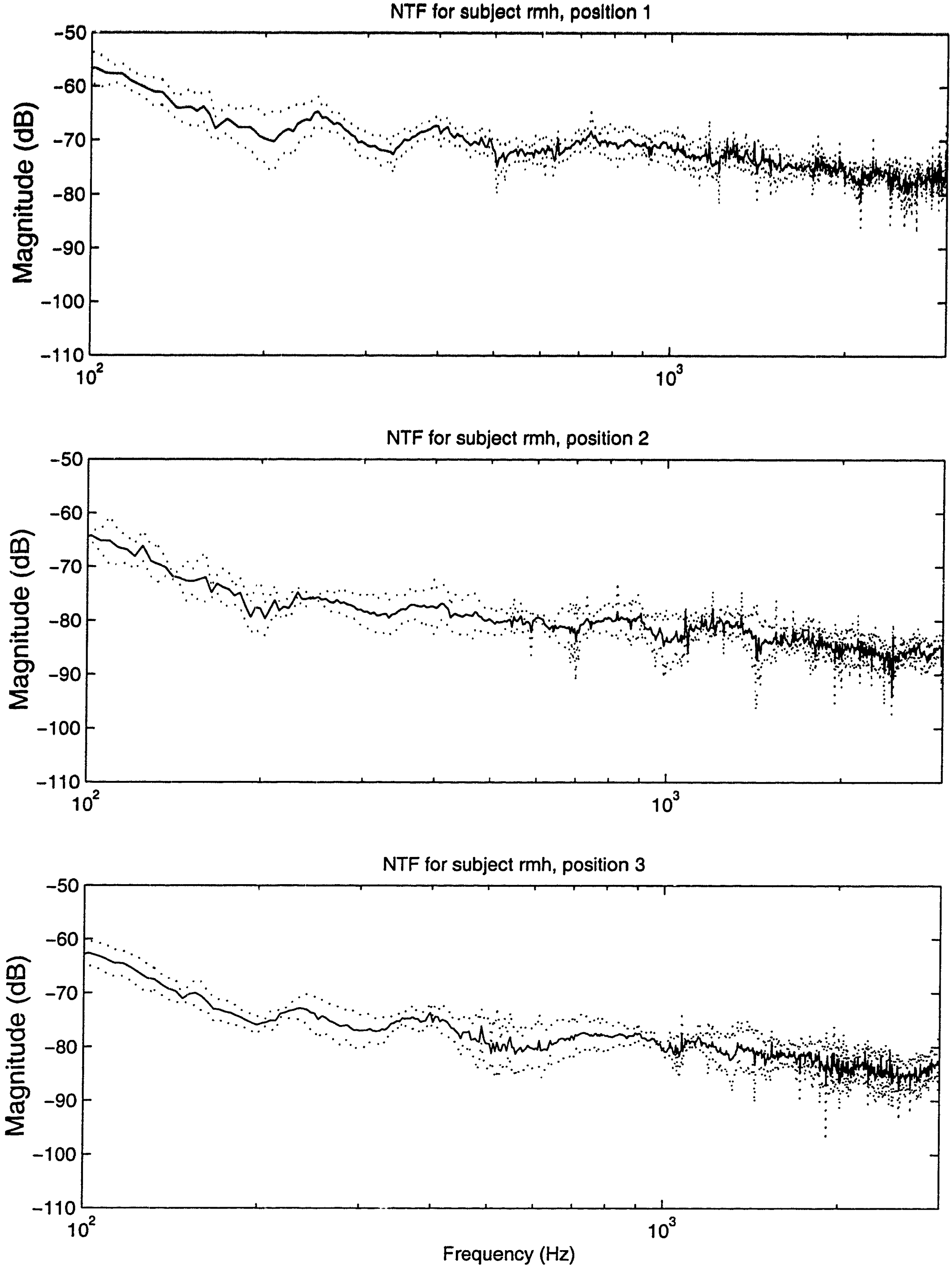


Figure A.13 The NTFs for subject rmh, a female laryngectomized subject.

VI. Appendix B.

This appendix contains the data for the 5 features extracted from the NTFs for each subject. The subjects are separated into the four subject groups discussed in Section III.

Subject	Position	Low f Slope (dB/octave)	Hi f Slope (dB/octave)
laryngectomized females			
rmh	1	-12	-5
rmh	2	-12	-5
rmh	3	-12	-4
kxb	1	-10	-5
kxb	2	-13	-5
kxb	3	-12	-5
lxh	1	-10	-6
lxh	3	-10	-6
mxm	1	-22	-6
mxm	2	-14	-6
mxm	3	-12	-5
laryngectomized males			
efp	1	-10	-4
efp	2	-12	-6
efp	3	-11	-5
jfs	1	-15	-7
jfs	2	-10	-5
jfs	3	-8	-5
rsm	1	-5.5	-4
rsm	2	-5	-5
rsm	3	0	-6
ssw	1	-10	-5
ssw	2	-8	-4
ssw	3	0	-6
reb	1	-9	-4
reb	2	-12	-5
reb	3	-12.5	-5
laryngeal males			
gsm	1	-8	-6
gsm	2	-8	-6
gsm	3	-8	-5
eag	1	-10	-5
eag	2	-8	-5
laryngeal females			
dyn	1	-10	-6
dyn	2	-10	-6
dyn	3	-9	-5
jls	1	-12	-4
jls	2	-10	-6
jls	3	-12	-5

Subject	Peak Freq (Hz)	Peak Gain (dB)	Max Attenuation (dB)
laryngectomized females			
rmh	101.6	-56.5	22.6
rmh	101.6	-64.2	24.2
rmh	105.5	-62.6	23.8
kxb	101.6	-59.9	20.9
kxb	101.6	-67.7	20.6
kxb	101.6	-68.9	22.8
lxh	97.7	-71.8	24.9
lxh	113.3	-78.4	29
mxm	101.6	-66.1	24.1
mxm	101.6	-65.9	25.4
mxm	101.6	-71.4	25.5
laryngectomized males			
efp	101.6	-64.8	-21.3
efp	105.5	-73.8	27.4
efp	101.6	-67.3	22
jfs	101.6	-62.4	26.4
jfs	105.5	-65.3	19.8
jfs	101.6	-73.8	21.2
rsm	117.2	-60.9	20
rsm	109.4	-60.4	21.7
rsm	113.3	-64.9	17.3
ssw	113.3	-74.2	17.9
ssw	125.1	-83.2	21.8
ssw	257.9	-68.8	29.1
reb	109.4	-70.1	25.3
reb	97.7	-62.8	27
reb	101.6	-71.3	27.8
laryngeal males			
gsm	105.5	-68.8	24.2
gsm	97.7	-66.4	22.8
gsm	97.7	-70	24.8
eaq	113.3	-63.7	21
eaq	109.4	-65.6	22.7
laryngeal females			
dyn	109.4	-71.4	27.3
dyn	109.4	-83.6	28.5
dyn	109.4	-74.7	27
jls	101.6	-66.1	24.6
jls	101.6	-80.4	26.3
jls	105.5	-71.1	24.9

VII. References

Atal, B. S., & Hanauer, S. L. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. J. Acoust. Soc. Am., 50 , 637-655.

Barney, H. L., Haworth, F. E., & Dunn, H. K. (1959). An experimental transistorized artificial larynx. In B. Weinberg (Ed.), Readings in Speech Following Total Laryngectomy . Baltimore: University Park Press.

Chari, V., Espy-Wilson, C., Huang, C., MacAuslan, J. and Walsh, M. (1998) Improvement of Electrolaryngeal Speech by Adaptive Filtering," J. Speech Lang. and Hear. Research, 41,1253-1264.

Fujimura, O., & Lindqvist, J. (1971). Sweep-Tone Measurements of Vocal Tract Characteristics. J. Acoust. Soc. of Am., 49 (2), 541-557.

Goldstein, E.A.(1998). An Improved Prosthetic ElectroLarynx Controlled by Pressure Variations and Electromyographic Signals. Final Project Report for Engineering Sciences 100hf, Harvard University.

Hillman, R.E., Walsh, M.J., Wolf, G.T., Fisher, S.G., & Hong, W.K. (1998) Functional Outcomes Following Treatment For Advanced Laryngeal Cancer. Ann. Otol., Rhino., and Laryngol, 107, 1-27.

House, A. S., & Stevens, K. N. (1958). Estimation of Formant Band Widths From Measurements of Transient Response of the Vocal Tract. J. Speech Hear Res., 1 (4), 309-315.

Houston, K.M. (1997) Report to Voice Project Group of the Keck Neural Prosthesis Center.

Kinsler, L.E., Frey, A.R., Coppens, A.B., & Sanders, J.V. (1982) Fundamentals of Acoustics. New York: John Wiley and Sons, Inc.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variation among female and male talkers. J. Acoust. Soc. Am., 87 (2) 820-857.

Norton, R. L., & Bernstein, R. S. (1993). Improved LAboratory Prototype ELectrolarynx (LAPEL): using inverse filtering of the frequency response function of the human throat. Ann Biomed Eng, 21 (2), 163-174.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of Vowels. J. Acoust. Soc. Am., 24 (1), 175-184.

Qi, Y., & Weinberg, B. (1991). Low-Frequency Energy Deficit in Electrolaryngeal Speech. J. Speech Hear Res., 34 , 125-1256.

Sisty, Nancy L., & Weinberg, Bernd. (1972). Formant Frequency Characteristics of Esophageal Speech. J. Speech and Hear Res., 15, 439-448.

Stevens, K.N. (1998) Acoustic Phonetics. Cambridge: MIT Press.

Weiss, M. S., Yeni-Komshian, G. H., & Heinz, J. M. (1979). Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx. J Acoust Soc. Am, 65 (5), 1298-1308.

Wodicka, G. R., Lam, A. M., Bhargava, V., & Sunkel, D. (1993). Acoustic impedance of the maternal abdomen. J. Acoust. Soc. Am., 94 (1), 13-18.

THESIS PROCESSING SLIP

FIXED FIELD: ill. ______ name ______

index ______ biblio ______

► COPIES: [✓ Archives] Aero Dewey [✓ Eng] Hum

Lindgren Music Rotch Science

TITLE VARIES: ► ☐ ______

NAME VARIES: ► ☑ Seth

IMPRINT: (COPYRIGHT) ______

► COLLATION: [illegible] 94 p.

► ADD: DEGREE: ______ ► DEPT.: ______

SUPERVISORS: ______

NOTES:

cat'r: date:

page:

► F41

► DEPT: E.E.

► YEAR: 1999 ► DEGREE: S.M.

► NAME: MELTZNER, Geoffrey S.