Chapter 6

Summary and Conclusion

This dissertation has resulted in an advanced model for the analysis and synthesis of pathological vowels. The success of the process is demonstrated in the fact that many of the synthetic vowels are indistinguishable for the original vowels from which they were modeled. The overall development proceeded as illustrated in Fig. 6.1, and is now summarized.

In regard to voice analysis as described in Chapter 2, the steps of the analysis phase of voice processing have been detailed. Using the source-filter model of voice production as a basis, voices have been parameterized into formants, LF source waveform, fundamental frequency time history, amplitude time history, FM nonperiodic effects, AM nonperiodic effects, and aspiration noise. The limitations and practical aspects of LP analysis for inverse filtering and the determination of formants have been described as well as source waveform fitting using the LF model. The nonperiodic features of pathological voices have been expressed in terms of AM and FM variations and aspiration noise. Gaussian distributions have been shown to model the small, high frequency effects in AM and FM variations. Aspiration noise was found to be well modeled with spectrally shaped Gaussian noise. The results of analysis form a set of parameters that may be used both for comparison of voices and for the generation of synthetic versions for perceptual testing.

In regards to external stimulation (ES) of the vocal tract, a major limitation of formant analysis and inverse filtering pathological voices is the reliance upon the (possibly spectrally deficient) source to reveal formant information. Segregation of the source waveform from the vocal tract using LP and inverse filtering may be ambiguous in the case of pathological voices, resulting in difficulty in achieving good vowel quality in synthesis. Without a good glottal source energy representation across the entire spectrum of interest (which is supplied mostly by the sharp return phase of a normal voice), resonances in pathological voices may not be detected with LP and other techniques. By externally stimulating the vocal tract with a spectrally rich source (chirp, impulse, white noise, etc), all resonances are clearly detected. Information obtained from ES analysis may then be combined with other approaches to yield more accurate formants and inverse filtered waveforms. The application of ES testing was successfully demonstrated via a progressive series of experiments described in Chapter 3. Formants of a known physical form (tube) were detected at the expected frequencies. Application of ES testing to the vocal tract revealed high-resolution detection of the expected formants. In addition, comparison of ES test results with traditional LP and FFT analysis of the same voices reveals ES testing produces higher resolution, more detail, and additional resonances not detected at all by LP and FFT. Experiments with simulated pathological (breathy) voices demonstrate a particular case in which ES testing improves formant estimation. Problems of articulator movement during ES testing may be solved by any of several technical approaches. Thus, the value of ES testing for pathological voice analysis is illustrated.

In regard to voice synthesis, synthesis provides a valuable tool in the study of pathological voices. Synthetic versions of pathological vowels measure levels and changes in pathological qualities in terms of their model parameters of synthesis, such as aspiration noise level, tremor, shimmer, and HFPV. Synthetic vowels may also be used to test perceptual significance of model parameters. Chapters 4 and 5 have described the efforts for re-synthesis of pathological vowels. The function and implementation of two synthesizers has been explained in Chapter 4. A hardware-based real-time synthesizer was constructed with the capability of generating immediate responses to changes in voice model parameters. The real-time synthesizer was designed with an expansible architecture allowing easy addition new features and model parameters, deterministic real-time performance, and the capability to perform real-time closed loop control by performing all system computations within a single sample period. A software synthesizer based on MATLAB was implemented to provide the capability to quickly code and evaluate complex algorithms that would be more time-consuming to code in real-time. The algorithms for implementing synthesis model parameters derived in analysis defined in Chapter 2 (LF source parameters, formants, aspiration noise, etc.) have been described in Chapter 5. Lastly, the validity of the overall analysis/synthesis process was tested by closing the loop with re-analysis of synthesizer outputs and with listener comparisons of original and synthetic vowels, as described in Chapter 5.

Contributions of this dissertation include:

1. Improved analysis techniques, especially those that measure NSR.

2. Improved AM and FM demodulation techniques which are effective for pathological voices.

3. Demonstration of the applicability of external source vocal tract transfer function testing for pathological voices.

4. Implementation of real-time and software-based voice synthesizers optimized for pathological vowels.

A key finding is that modeling of nonperiodic pathological voices with AM, FM, and aspiration noise does improve the accuracy of analysis and fidelity of synthesis, which was the original experimental question posed.

6.1 Future Work

The current approach to analysis/synthesis yields voices approach the originals in many cases, but some limitations were observed, which suggest possible future areas of work:

1. The current synthesizer uses a constant value for the level of aspiration noise (relative to voiced energy). Some pathological voices are particularly unsteady (even though the subjects were asked to maintain constant volume), and they exhibit considerable shift in the ratio of unvoiced and voiced energy during the one-second sample. Making the aspiration noise level a function of time will improve the synthetic versions of these voices.

2. The current synthesizer uses constants for the formants for the entire sample. FM demodulation should remove all apparent fundamental frequency variation. However, when listening to the FM demodulated natural voice time series, the impression of fundamental frequency variation can still be perceived in some cases despite the fact that reanalysis of this demodulated natural voice shows constant fundamental frequency. A possible explanation of this effect could be formant modulation. Inclusion of time varying formants in analysis and synthesis should eliminate this effect and improve the fidelity of synthesis. Efforts at implementation are currently under way.

3. Careful comparison of some of the original and synthetic voice samples in some cases reveals short period differences in quality difficult to assign to any parameter. The current model uses constant LF parameters for the entire segment. Analogously to formants, time varying LF parameters may improve fidelity and may also explain remaining small differences between measured and perceived aspiration noise levels.