Proceedings of Meetings on Acoustics

Volume 26 http://acousticalsociety.org/

171st Meeting of the Acoustical Society of America

Salt Lake City, Utah

23-27 May 2016

Psychological and Physiological Acoustics: Paper 1aPP48

Spectral-normalization filter for subjective analysis of the aging voice

Mark L. Berardi and Eric J. Hunter
Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, 48824; mberardi@msu.edu; ejhunter@msu.edu

Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84604; kentgee@byu.edu

Voice quality changes with age. In many cases, these voice changes result in a lower quality of life. Because one way of identifying these voice quality changes is through perceptually estimating talker age, correlations made between estimated talker age and acoustic analysis can provide insight to the possible physiological degeneration related to vocal function. While most perceptual studies investigating estimated talker age are cross-sectional, a longitudinal study of single speakers could provide additional details in the progressive degeneration of the voice quality. Nevertheless, one limitation of these studies is that perceptual ratings of voice quality or talker age in a longitudinal study could be biased by recording quality. Further, the spectral qualities of recordings from earlier decades are limited by the technology used. In this paper, a spectral-normalization filter was developed and applied to a corpus of recordings from an individual spanning about 50 years (1959 - 2007) to reduce this impact of these limitations. The filter was shown to be effective in normalizing the autospectra of the recordings and the fundamental frequency was unaffected by the filter. Preliminary subjective analysis suggests that the recording quality of all the files were perceptually similar.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0000770

1. INTRODUCTION

Laryngeal tissues and voice quality change with age, often resulting in a lower quality-of-life. One study showed that more than 50% of the surveyed members of the geriatric population had impaired quality-of-life as a result of their voice (Golub et al. 2006). Since perceptual estimations of talker age is one way of identifying voice quality changes, correlations between estimated talker age and acoustical analysis may provide insights into possible physiological degeneration related to vocal function. Although studies investigating estimated talker age have been traditionally cross-sectional in design, a longitudinal study can provide additional key details related to the progressive degeneration of voice quality (Hunter et al. 2012). However, a longitudinal sample of recordings used in perceptual ratings of voice quality or talker age could be biased by improvements in recording quality as recording equipment has evolved. Therefore, an approach to offset this limitation needs to be developed. In this study, a spectral-normalization filter was developed and applied to a corpus of recordings from an individual spanning nearly 50 years (1959 - 2007).

2. EXPERIMENTAL METHODS

Spectral-normalization filters were developed to remove recording bias in subjective estimation of talker age. A comparison of the autospectra of the filtered recordings was used to validate the normalization. Additionally, conventional speech parameters were computed and compared across the different filters to investigate the effects of the filter on acoustic measures of dysphonia.

Prior to applying the spectral-normalization filter, all the recordings were amplitude normalized and bandpass filtered between 80 Hz and 5 kHz. The filters were designed in MATLAB® to shift spectral amplitudes of individual recordings relative to a reference spectrum. The average spectrum from recordings from 1959 was used as the reference recording. The procedure for spectral normalization is as follows:

a) The autospectra of the reference recordings were averaged to obtain the reference autospectrum, G_{ref}. The autospectrum calculation used 1024 samples per block.
b) An amplitude transfer function between each of the recordings from 1959 to 2007 and the reference autospectrum was obtained by $H(f) = \sqrt{G(f)/G_{ref}(f)}$, where $H(f)$ is the transfer function and G is the autospectrum of the recording being analyzed. The transfer function was made two-sided and interpolated to be the same length as the individual record being normalized.
c) The Fourier spectrum, $g(f)$, of the recording to be normalized was obtained using MATLAB's function, fft.m.
d) A normalized Fourier spectrum for the recording was calculated as $g_n(f) = g(f)H(f)$.
e) Finally, a new wave file was written using the inverse-fast-Fourier transform of the spectrum, $y_n(t) = ifft(g_n(f))$.

3. RESULTS

The filtering method was validated through changes in the autospectrum resulting in normalized or amplitude-similar autospectra. Figure 1 shows the reference spectrum (G_{ref}). Figure 2 shows the autospectra for the original, filtered segmented recordings (filter relative to the average of all recordings), as well as the reference autospectrum, G_{ref}, for the 2007 recordings. In this Figure, it was observed that the spectral shape remained consistent after the filtering, although the amplitudes were adjusted so that they were more similar to the target filter.

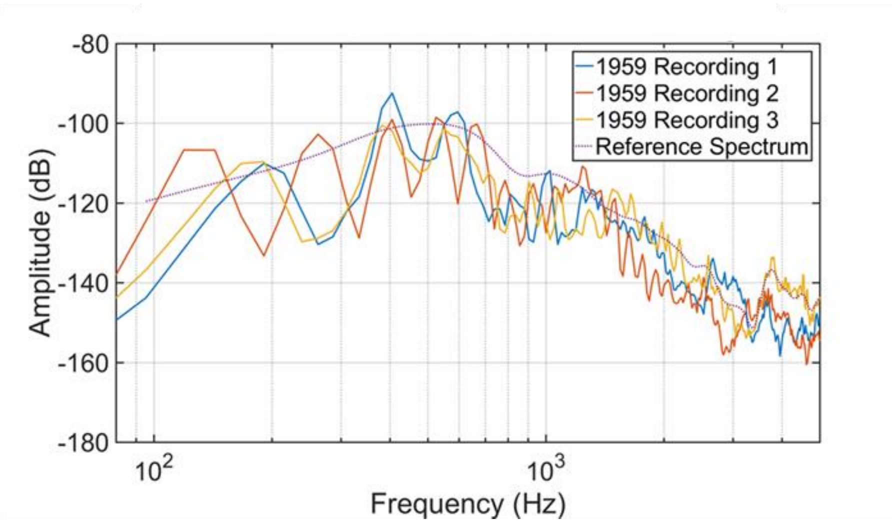

Figure 1. Reference spectrum calculated from the recordings from 1959.

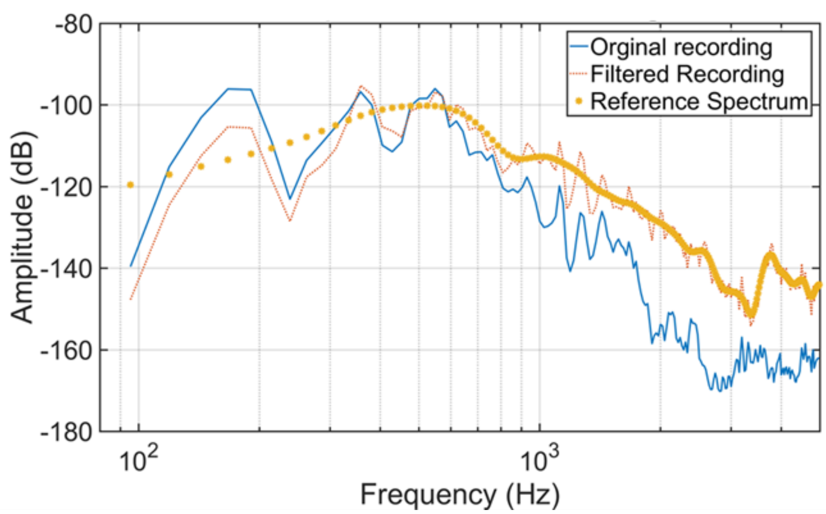

Figure 2. Autospectra of the original, filtered, and reference recordings from 2007.

An investigation of the speech-acoustic parameters shows that the core temporal speech parameters remained unchanged. One example is in fundamental frequency. Figure 3 illustrates that the fundamental frequency of the original segmented recordings and the three filters are nearly identical. Further analysis showed that a linear fit of the fundamental frequency plots had a positive slope, which is consistent with previous work by Hunter et al. (2012).

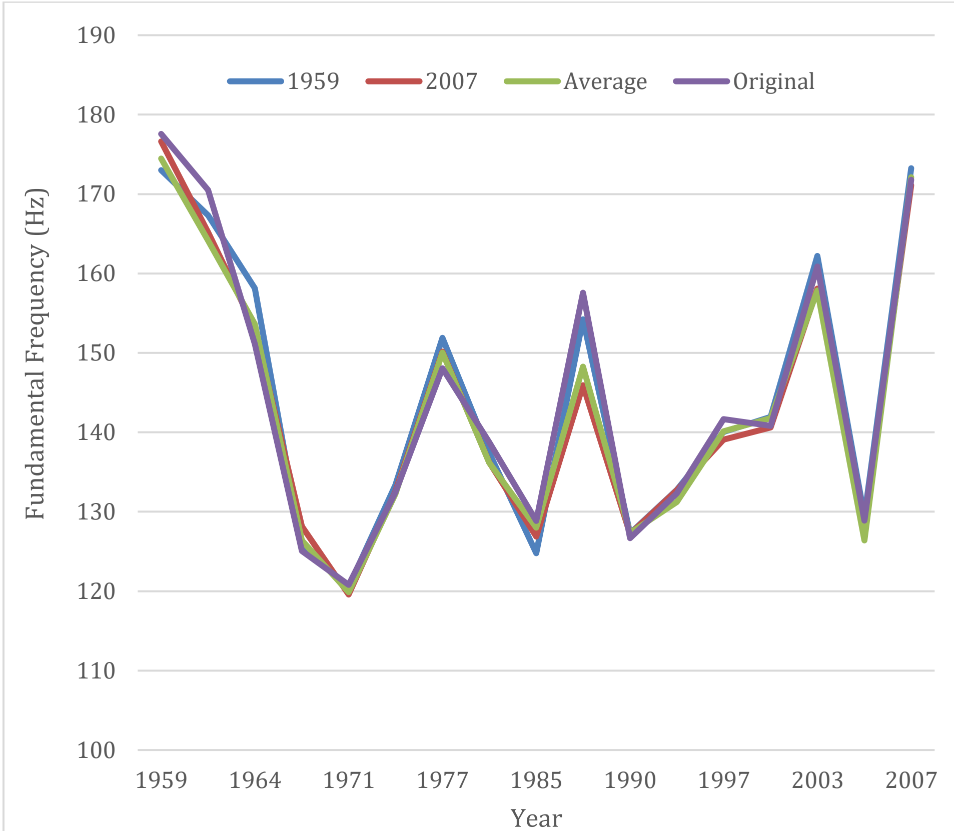

Figure 3. Fundamental frequency over 50 years with three different spectral-normalization filters and the original segmented recording.

Preliminary subjective analysis by the research team suggests that the filters were effective in normalizing the effects of recording equipment. It was found that the filter referenced to 1959 had the least amount of artificial noise.

4. CONCLUSIONS

While longitudinal sample of recordings provide key insights into how a voice changes over time, improvements in recording quality over time might affect the results. Therefore, the current study developed spectral-normalization filters to remove such recording bias. As seen by a comparison of the spectra, the developed filter was effective in amplitude normalizing the spectrums of the recordings used for subjective estimation of talker age. Speech-acoustic analysis resulted in changes of spectral character, but not periodicity character (as seen in a consistent fundamental frequency). This suggests that filters did not change core characteristics of the speech sound but accounted for spectral differences in recording equipment.

Preliminary subjective analysis suggests that the recording quality of all the files were perceptually similar. The filtered recordings will be implemented in a future more exhaustive study estimating talker age. One limitation of the current study is that the filtered recordings still have differences in background noise and reverberation. The filter also added additional artificial noise because of the signal processing. Further work will try to account for this added noise.

REFERENCES

Golub, J. S., et al. (2006). "Prevalence of perceived dysphonia in a geriatric population." Journal of the American Geriatrics Society, **54**(11): 1736-1739.

Hunter, E. J., et al. (2012). "Age and Speech Production: A 50-Year Longitudinal Study." Journal of the American Geriatrics Society, **60**(6): 1175-1177.

Hunter, E. J., et al. (2015). "Listener estimations of talker age: A meta-analysis of the literature." Logopedics Phoniatrics Vocology, **41**(3): 101-5.