NOISE AND APERIODICITY IN THE GLOTTAL SOURCE: A STUDY OF SINGER VOICES

Perry R. Cook (PRC@CCRMA.STANFORD.EDU)

Center for Computer Research in Music and Acoustics (CCRMA), Stanford Univ.

ABSTRACT

Methods for extracting, analyzing, and visualizing the time-domain nature of aperiodicities in the glottal source are presented. A study of noise in the voice was conducted using 12 trained singers. Average Normalized Noise Power (NNP) was computed as a function of pitch and volume. The time domain nature of the noise signal is investigated and compared to theoretical predictions from the principles of fluid dynamics.

1. INTRODUCTION

The quasi-periodic oscillations of the glottis exhibit small period-to-period deviations in the waveform [8], much of which is brought about by bursts of noise in the oscillator itself. This paper investigates aperiodicities in steady state sung tones, using new techniques for extraction, visualization, and quantification.

2. EXTRACTION TECHNIQUES

To study the aperiodic component of a quasi-periodic signal, some method must be used to identify and separate the periodic and aperiodic components. The two methods used in this study operate in the time domain, and yield slightly different results because of the definition of periodicity each assumes. The first involves using a least-squares periodic predictor to yield an error signal representing the component of the signal which is not periodic. The second method uses period similarity processing [10], with the added improvement of sinc-interpolated sampling rate conversion [9]. Each period of a steady state vowel is compared to all others and resampled to yield a least-squares difference. All such resampled periods are then averaged as vectors to yield a prototype period, which contains no noise component in the limit as the number of periods approaches infinity. The prototype is subtracted from each period to yield the period residuals. This method has the advantage that the time domain connection between the signal and residual periods is preserved. Figure 1 shows the waveforms of a male singing the vowel /ʌ/ (bug) at 100 Hz., the periodic component, and the residual.

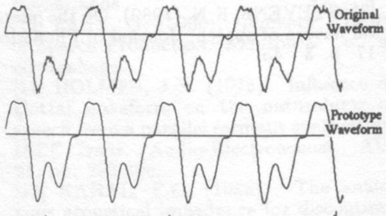

Figure 1. Original (top), prototype (periodic), and amplified residual of male vowel /ʌ/ sung at 100 Hz.

3. VISUALIZATION TECHNIQUES

Period synchronous methods of analyzing the extracted noise residual signals were used. The techniques involve identifying the periods of the original signal (already accomplished if extraction is done by the period similarity method). Identification of the periods involves detection of some time domain feature using a method such as low-pass filtering and zero-crossing detection. Each detected period yields a pointer into the time-domain residual signal, and thus the noise 'periods' can be subdivided (into 6 sections for this study) to inspect the behavior.

In the period-synchronous noise power analysis method, the noise power (sum of squares) is computed in each of the sub-sections. These powers can then be plotted in three dimensions, with height representing power, one axis representing the period number, and one axis representing the position within the period. Inspection of this noise period power surface will show clear ridges and valleys (running in the direction of the period number axis) if the signal contains pulsed noise. The duty cycle can be deduced from the width of the ridges and valleys, and the dynamic range of the noise can be deduced from the ratio of the heights of the ridges and valleys. Averaging across the power surface in the direction of the ridges and valleys yields an estimate of the average noise power at particular times within a typical period. Figure 2 shows the noise power surface of the vowel of Figure 1. The placement of the ridges and valleys shows that noise is more likely to occur as the glottal folds open and close, and less likely when the glottis is completely closed or completely open.

Another analysis technique involves performing Discrete Fourier Transforms (DFTs) on each of the residual period subdivisions. Each DFT can be used to compute a power spectrum, and the power spectra corresponding to a particular period position can be averaged across all periods, yielding an

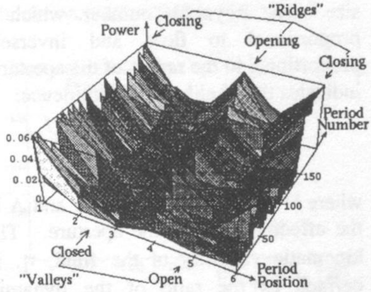

Figure 2. Noise period power surface of male vowel /ʌ/ sung at 100 Hz. Opening and closing phases of the glottis are noted.

estimate of the power spectrum of the noise at that position within a typical period. These spectra can be plotted in 3D as intensity (height) versus period position and frequency. Figure 3 shows the period spectral surface of the male vocal tone of Figure 1. The spectral energy clearly shifts toward higher frequencies at the instant where the glottal folds are opening.

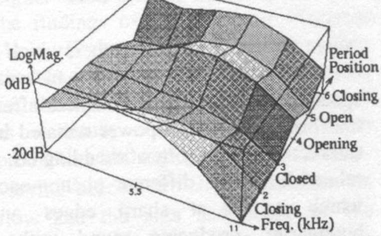

Figure 3. Noise period spectral surface of male vowel /ʌ/ sung at 100 Hz. Opening and closing phases of the glottis are noted.

4. GLOTTAL FLUID DYNAMICS

The passage of air at sufficient velocity through an aperture causes turbulent streaming and noise is generated. The flow is zero when the time varying aperture is closed, and the turbulence ceases if the aperture opens sufficiently or the flow decreases. The basic fluid dynamic equations quantifying turbulent jet formation and noise radiation are expressed in terms of flow and aperture size. The Reynolds number, which is proportional to flow and inversely proportional to the radius of the aperture, indicates the likelihood of turbulence:

$$Re = (2U) / (\eta (A \pi)^{1/2})$$

where U is the volumetric flow and A is the effective area of the aperture. The kinematic viscosity of the fluid, η, is defined as the ratio of the dynamic viscosity to the density, and is about 0.15 cm^2 per second for dry air. Turbulent streaming is likely if the Reynolds number is greater than a critical quantity, Re_{crit}, which is about 1,000 for a rectangular slit. If turbulence is present, noise is generated with a power proportional to V^8, or proportional to $(U/A)^8$ as expressed in terms of volume flow and area. The center frequency of the principal peak in the spectrum of the turbulent noise is given by:

$$f = (S V) / d = (S U \pi^{1/2}) / (2 A^{3/2})$$

where S is the Strouhal number, which is 0.15 for the center frequency of noise spectral density. Tube resonances affect the formation of, and power radiated by turbulent jets. Vortex shedding is a related but quite different phenomenon which occurs at sharp edges and boundaries, producing sound with a power which depends on lower powers of the flow-to-area ratio. Hirschberg [3] gives power relationships of $(U/A)^4$ and $(U/A)^6$ for turbulent sound radiation in a tube, and vortex dipole sound radiation in a tube, respectively.

Figure 4 shows the characteristics of a typical cycle of oscillation of the glottal folds. Views a) and b) are drawings of the superior and cross sectional views of the glottal folds. Graphs c) and d) show the effective area and volumetric flow (flow glottograph). Assuming a maximum flow of 300 cm^3 per second, and maximum dimensions of 2.6 mm by 11 mm for the vocal folds, the Reynold's number, radiated power, and center frequency of the radiated noise power are graphed as e) f) and g) of Figure 4.

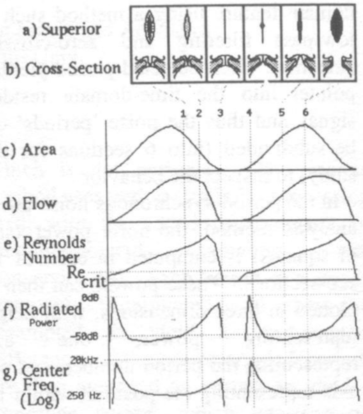

Figure 4. Phases of glottal oscillation, with area, flow, Reynolds number, radiated noise power, and noise spectrum center frequency.

The simple analysis of Figure 4 assumes that turbulence is instantly born when the dimensions and flow quantity are suitable, and the disturbance dies as quickly. A more detailed analysis of the behavior of pulsed turbulence was done by Kingston [6]. These studies investigated the effects of turbulent jets in tubes driven by pulsating sources of flow. The ratio of normalized pulsation frequency, Ω, to the Reynolds number was identified as an important measure of turbulent behavior. Ω is defined by:

$$\Omega = r^2 \omega / \eta = 2AF_0 / \eta$$

where r is the radius of the tube, A is the tube cross-sectional area, and F_0 is the frequency of phonation in Hz.

Given the large frequency range of the singing voice, and allowing for large deviations in tube cross-sectional area depending on the vowel, Ω can range from 100 at 50 Hz. in an /μ/ vowel, to 10^5 at 2000 Hz. in an /a/ vowel.

Three regimes of pulse-turbulence interaction were observed by Kingston, corresponding to high, medium, and low ratios. For low pulsation frequencies ($\Omega / Re < 0.04$), the flow is quasi-steady and follows the behavior indicated by the analysis of Figure 4. For high pulsation frequencies ($\Omega / Re > 0.1$), the turbulence is steady and independent of flow pulsations. For intermediate frequencies, the relation between turbulence and pulsation is complex, and is characterized by vortex resonance phenomena.

The average value of the Reynolds number from Figure 4 is 2750. Assuming a minimum vocal tract tube area of 0.15 cm^2, the transition region from pulse-turbulence interaction to non-interaction lies between 55 and 140 Hz. The maximum Reynolds number is 5860, yielding a transition region bounded by 120 and 300 Hz. From these calculations, pulsed turbulence is possible at phonation frequencies below 200 Hz. Even allowing for large deviations in the assumed parameters of flow and glottal area, it is still expected that low notes sung by bass singers might exhibit pulses, or perhaps dual pulses. As simulated by Iijima, Miki, and Nagai [5], periodic vortex shedding is also expected in the glottal region.

5. NOISE IN SINGER VOICES

Twelve highly trained singers (Bel Canto) were asked to produce sung tones at three dynamic levels without vibrato on the the neutral vowel /ʌ/ across their entire comfortable singing range.

5.1. Average Noise

Figure 5 shows the average Normalized Noise Power (NNP) of all singers as a function of phonation pitch. The least-squares fit through the points of an exponential function of the form $k_1 * f^{k_2}$ is $59 * f^{-1.2}$, indicating a relationship between noise power and phonation pitch which is close to 1/f.

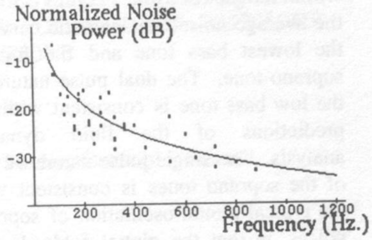

Figure 5. Normalized Noise Power (NNP) of the sung tones of 12 singers, graphed as a function of phonation pitch.

Given the fluid dynamic predictions that radiated noise power varies as a high power of flow, it may seem contradictory that NNP in singer voices was largely independent of dynamic level, and inversely proportional to frequency. A study of airflow in singer voices [7] found that flow increases slightly with both increasing pitch and loudness, but often airflow decreases in higher tones. This is also consistent with the findings of Cavagna and Margaria [1]. Higher tones are very often produced with a more 'pressed' voice, and the overall glottal resistance changes as a result. The nature of noise production in the glottis is that of a time-varying process which is dependent on flow and the area and shape of the aperture, so it is likely that any increase in flow is offset by changes in the time-varying area function. In the falsetto register there is a direct relationship between phonation frequency and flow [4], so there is a likelihood of higher noise power for increasing frequency in this range. All of the male test subjects showed an increase in noise power when entering the falsetto register, and most of the falsetto tones exhibited an increase in noise power with increasing frequency.

5.2. Pulsed Noise

The noise period power surfaces were

calculated for all singers, and averages were taken across the data at six points within the glottal cycle. Figure 6 shows the average noise characteristic curve of the lowest bass tone and the highest soprano tone. The dual pulse nature of the low bass tone is consistent with the predictions of the fluid dynamic analysis. The single pulse nature of two of the soprano tones is consistent with the typical glottal oscillation of soprano voices, in that the glottal folds do not close entirely. This is easily seen by the fact that the principal peak in the noise curve occurs when the glottal folds are in the 'closed' position.

All bass subjects exhibited dual pulse behavior in the low register, and on some tones in the high register. The alto and tenor subjects exhibited a shift from dual pulse at the low register to single pulse noise behavior in the high register. The Noise Dynamic Range (NDR), defined as the ratio of the highest noise power sample to the lowest, decreased weakly with pitch from about 10 dB in the lowest tones to 1 dB in the highest. No clear behavior as related to dynamic level was evident in any of the voices studied.

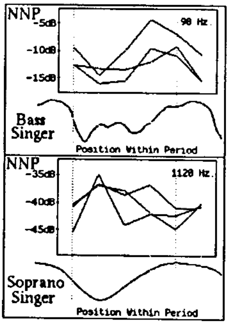

Figure 6. Normalized Noise Power (NNP) as function of period position for lowest bass and highest soprano tones, for three dynamic levels.

10. REFERENCES

[1] Cavagna, G. A. and R. Margaria (1968), "Airflow Rates and Efficiency Changes During Phonation," *Annals of the New York Academy of Sciences*, 155, 152-164.

[2] Cook, P. R. (1991), "Identification and Control of Parameters in an Articulatory Vocal Tract Model, With Applications to the Synthesis of Singing," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford Univ.

[3] Hirschberg, A. (1990), "Some Fluid Dynamics Aspects of Speech," *Fourth Colloq. Signaalanalyse en Spraak, COLSAS*, Instituut voor Perceptie Onderzoek, Eindhoven.

[4] Isshiki, N. (1965), "Regulatory Mechanism of Voice Intensity Variation," *J. Sp. Hearing Res.*, 7, 17-29.

[5] Iijima, H., N. Miki, and N. Nagai (1988), "Viscous Flow Analyses of the Glottal Model Using a Finite Element Method," *2nd Mtg. ASA & ASJ*, NN10.

[6] Kingston, G. C. (1975), "Experimental and Theoretical Studies of Pulsating Turbulent Flow," Ph.D. Dissertation, Clarkson College of Technology.

[7] Rubin, H. J., M. LeCover and W. Vennard (1967), "Vocal Intensity, Subglottic Pressure and Air Flow Relationships in Singers," *Folia Phon.*, 19, 393-413.

[8] Schumacher, R. T. and C. D. Chafe, (1990) "Detection of Aperiodicity in Nearly Periodic Signals," *Proc. IEEE Conf. ASSP*, Albuquerque, NM.

[9] Smith, J. O. and P. Gossett (1984), "A Flexible Sampling-Rate Conversion Method," *Proc. IEEE Conf. ASSP*, San Diego, CA, March.

[10] Yumoto, E., Y. Sasaki, and H. Okamura (1984), "Harmonics-to-Noise Ratio and Psychophysical Measurements of the Degree of Hoarseness," *J. Sp. Hearing Res.*, 27, No. 2-6.