15th Triennial World Congress, Barcelona, Spain

LINEAR PREDICTION ACOUSTICAL MODELLING OF FREE FIELD COUGH SOUND

A. Van Hirtum [*], **K. Janssens** [*], **K. Demuynck** [**], **D. Van Compernolle** [**], **D. Berckmans** [*]

[*] *Department of Agro-Engineering and Economics, K.U.Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium, daniel.berckmans@agr.kuleuven.ac.be*

[**] *Department Electrical Engineering, K.U.Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium*

Abstract: In this paper the performance and assumptions of linear prediction acoustical modelling are assessed on the free field cough sound. Four distinct free field cough classes originating from animal and human species in different health conditions are considered. For each cough class the vocal tract formants are estimated from the linear prediction parameters.

Keywords: Signal processing, Spectral analysis, Linear prediction, Autoregressive modelling, Signal-to-noise ratio, Biomedical system

1. INTRODUCTION

As reflex-generated perturbation of the respiratory function cough is an important symptom in many respiratory diseases or irritations (Irwin, *et al.*, 1977). The simple, noninvasive, nonhazardous, contactless and inexpensive nature of acquiring information of the respiratory system using free field cough-sound registration makes it an attractive candidate for on-line follow-up and clinical diagnosis (Korpas, *et al.*, 1996). However the application of this method is still limited due to the inability to extract adequate objective information and the lack of a full understanding of the origin of the cough sound. Auditive characterization of the cough sound resulted in several common labels as brassing, barking, whooping, etc. A more objective cough-description is obtained by time-frequency analysis (Korpas, *et al.*, 1996; Murata, *et al.*, 1996). Modelling of the free field cough acoustic waveform aims to parameterize the cough sound signal for analysis and physical interpretation. Model parameters are found by performing a time or frequency match between the original signal and that generated by the model. In this paper linear prediction acoustical modelling of the free field cough waveform is assessed.

2. MATERIALS AND METHODS

2.1 *Acoustical data: animal and human studies*

In this study 'acute' cough due to a common cold and 'voluntary' cough on request are registered on respectively 3 suffering and 9 healthy non-smoking human individual subjects between 20 and 30. 'Chemical' and 'chronic' cough are registered during reproducible eliciting of coughing on individual Belgian Landrace piglets, aged 9 weeks, by respectively nebulization of citric acid (2 piglets) (Van Hirtum, *et al.*, 2000; Moreaux, *et al.*, 1999) and inoculation of a respiratory infection (2 piglets) (bronchopneumonia with Pasteurella multocida) (Kobisch, *et al.*, 1996), without disturbing the animal's behaviour. This resulted in respectively 48 'acute', 36 'voluntary', 1883 'chronic' and 119 'chemical'-induced cough samples. Free field acoustic registration at 22050 Hz is performed with a standard multi-media

microphone (20-20kHz frequency response) and sound-card (16 bit). The microphone was positioned at a distance of respectively 0.3 up to 0.5 and 0.3 up to 1.7 m from the human or piglet subject. So acoustical studies on both species resulted in 4 cough-classes.

2.2 *Linear prediction for spectral analysis*

Linear prediction (LP) time-domain acoustic modelling is commonly applied in speech processing using a source-filter arrangement to model the vocal tract system (Rabiner, *et al.*, 1993). In general it is assumed that the source is located at the glottis and that a linear filter is adequate to model the frequency properties of the vocal tract. Furthermore for analysis it is assumed that no information about the excitation of the vocal tract is known and that the sound waveform can only be modelled from its previous values. Based on both assumptions the linear vocal tract filter defines an autoregressive (AR) model of the signal, in which the current sample, y(t), is predicted from a linear combination of a finite number (n_a) of past samples

$$\hat{y}_p(t) = -\sum_{k=1}^{n_a} a_k y(t-k) \qquad (1)$$

where $\hat{y}_p(t)$ is the predicted signal sample and n_a the prediction order. The coefficients a_k are assumed constant over the analysis frame. The prediction error or residual, $e_p(t) = y(t) - \hat{y}_p(t)$, represents either structure in the sound waveform which is not captured by the model or randomness which can inherently not be modelled. The LP spectral estimate only makes sense if the model is correct (inclusive stationarity of the vocal tract, white noise or null source), if not it will make the source look as much as white noise as possible, putting all spectral shaping into the transfer function. For a good model, the residual has no predictable structure and appears as white noise. The vocal tract transfer function is expressed in the z-domain as the all-pole system:

$$\hat{H}(z) = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^{n_a} a_k z^{-k}} \qquad (2)$$

with G the gain term. The estimated vocal tract transfer function, $\hat{H}(z)$, represents the combined effects of the glottal wave shape, vocal tract response and lip radiation. A match between the spectral envelope of the waveform and the frequency response of $\hat{H}(z)$ is obtained if the parameters, a_k, are derived to minimize the mean squared prediction error, E, over the analysis frame length, which leads to maximum likelihood estimates of parameters assuming prediction errors have Gaussian distributions. Due to the spectral matching property of the mean-squared error criterion, linear prediction analysis can be used to obtain a smoothed estimate of the short-time spectral envelope of the sound waveform. Estimates of the formants are obtained by locating peaks in the smoothed spectral envelope or by factorizing A(z) into its constituent poles. Each formant is approximated by a complex-conjugate pole pair, $[p_i, \bar{p}_i]$ with $p_i = r_i exp(j\psi_i)$, which forms a second order filter with transfer function $A_i(z)$, given by

$$A_i(z) = 1 + a_1 z^{-1} + a_2 z^{-2} \qquad (3)$$

The frequency of the formant, F_i, is determined from the pole angle and the bandwidth, B_i, from the radius.

$$F_i = \psi_i / 2\pi T \qquad (4)$$

$$B_i = -(1/\pi T) log|r_i| \qquad (5)$$

The spectrum is unique in the range $-f_s/2 < f < f_s/2$ and repeats at multiples of the sampling frequency, $f_s = 1/T$. The transfer function, $\hat{H}(z)$, is stable if all the poles lie inside the unit circle. Model performance is objectively evaluated by the prediction signal-to-noise ratio (SNR):

$$prediction\, SNR = 10\, log_{10} \left(\frac{\sum_{t=0}^{N-1} y^2(t)}{\sum_{t=0}^{N-1} e_p^2(t)} \right) (6)$$

3. RESULT AND DISCUSSION

3.1 *Effect of waveform pre-processing*

In accordance with previous research, (Van Hirtum, *et al.*, 2001), 100 Hz high pass filtering improves the signal-to-noise ratio by eliminating low-frequency noise (mainly from mechanical origin such as ventilation noise) while preserving the intended acoustical waveform at its maximum. The strived $\hat{H}(z)$ is the estimated vocal tract transfer function. However in section 2.2 it is pointed that $\hat{H}(z)$ represents the combined effects of the glottal wave shape, vocal tract response and lip radiation. Assuming that the glottal volume velocity can be approximated by a two-pole model (-12dB/octave) and the lip radiation by a single zero (+6B/otave), the combined effect of glottal wave shape and lip radiation introduces a -6dB/octave shift to the magnitude spectrum (Deller, *et al.*, 1993) and can be approximated by a single pole P(z).

$$P(z) = 1 - \mu z^{-1},\ \mu = 0.95 \qquad (7)$$

Therefore pre-emphasis with the first-order filter P(z) is commonly applied to the waveform in

order to remove this trend and consequently to spectrally flatten the waveform by respectively suppressing and enhancing lower ($< 4kHz$) and higher ($> 4kHz$) frequency components. In order to minimize the waveform discontinuities at the beginning and ending of the analysis frame it is common to (Hanning) window each analysis frame so as to taper frame begin- end ending to zero. The effect of these three frequently used pre-processing steps (noise-filtering, pre-emphasis and windowing) and there mutual combinations on the estimated vocal tract transfer function, $\hat{H}(z)$, is assessed on the experimental data by comparison of the obtained prediction SNR given in Equation 6. It appears that for all experimental data classes pre-emphasis yields the largest prediction SNR followed by firstly no pre-preprocessing and secondly successively high-pass filtering and windowing. Therefore in the following it is chosen to pre-process the waveform by pre-emphasizing.

3.2 *LP model order*

In general the model order is determined by the properties of the data, the specific form of H and the form of the vocal tract excitation. In this paper the model form of H is fixed as a LP model indicated in Equation 2 because of its simple closed-form solution, complete separation of the source and the vocal tract filter in synthesis and a possible direct interpretation in terms of a loss-less acoustic tube model of the vocal tract (Rabiner, *et al.*, 1993). No assumptions are made concerning the form of the vocal tract excitation and the properties of the data are optimized for modelling during the pre-processing described in subsection 3.1. For speech sampled at 8kHz typical analysis orders range from $n_a = 10 - 16$ in correspondence with 4 formants requiring a minimum of 8 poles (4 pole pairs) with some poles added to count for the effect of glottal shaping, lip radiation and nasal coupling. Bearing this in mind the LP model is applied to the 4 classes of cough data described in subsection 2.1 with the model order respectively equal to 5, 8, 10, 12, 14, 15, 16, 18, 20 and 25. The analysis frame-length is set to 45 msec and is shifted with one third of the frame-length or 15 msec. The effect of varying the model order on the mean prediction SNR and associated standard deviation for each of the 4 cough-classes is presented in Table 1. As expected from the LP spectral matching property the prediction SNR in Table 1 increases with increasing model order. However this property might easily introduce over-parameterization of the waveform under study. Comparison of waveform and error spectra from LP with variation of model order for a representative cough of (a) animal: chronic, (b) animal: chemical, (c) human: acute and (d) human: voluntary are shown in Figure 1. The green, magenta and red fit on top of the blue waveform spectra corresponds respectively with $n_a = 5$, $n_a = 10$ and $n_a = 14$. Based firstly on the improvement in mean prediction SNR by increasing the order with one step in Table 1 and secondly on the visual inspection of smoothed spectral matching of the harmonic resonances as illustrated in Figure 1 the model order was set to $n_a = 14$ for the 4 cough classes. Other subjective criteria like auditive interpretation of the synthesized cough sound or objective criterion functions incorporating the variance on the parameters as among others defined in Akaike's Information Criterion (AIC), Finite sample Information Criterion (FIC) or Young Identification Criterion (YIC) are not applied. Although such objective criteria might improve the LP modelling here it is assumed that the parameters are not biased due to the de-noising incorporated in the pre-processing described in subsection 3.1 and the fairly good estimation of the harmonic resonances as shown in the plots of Figure 1. The poor modelling of the valleys between the resonance peaks is inherent to LP modelling spectral matching since lower spectral values contribute less to the mean-squared error criterion and therefore are less accurately modelled. For the animal species the class resulting from infected subjects (animal: chronic) exhibits the best model performance, while for human subjects the class resulting from healthy subjects (human: voluntary) shows a slightly better performance compared to human: acute. For infected subjects the performance of the LP model on animal subjects (animal: chronic) exceeds the performance on the human subjects (human: acute), while the opposite holds for the healthy classes (animal: chemical and human: voluntary). The cough-class due to chemical irritation (animal: chemical) clearly shows the worst model performance. The high and fairly constant standard deviations over all model orders in Table 1 indicate a variable nature of the cough waveform for all cough-classes. This finding and the overall low SNR-values in Table 1 cast doubts on the validity of the assumption of a two-pole model for the glottal volume velocity during sound production. Therefore other model structures need to be assessed.

3.3 *Estimation of LP model parameters*

The mean LP parameter estimates a_k with associated relative standard deviations $\xi(a_k)$ (in %) obtained with the model order fixed at $n_a = 14$ as described in subsection 3.2 for all 4 cough-classes is given in Table 2. The large relative standard deviations $\xi(a_k)$ on the mean parameter estimates for $n_a = 14$ again indicate a variable nature of the cough waveform for all cough-classes. Lower

Table 1. Effect of variation of LP model order on mean prediction SNR (mean) and associated standard deviation (std).

Model order	animal: chemical		animal: chronic		human: voluntary		human: acute	
n_a	mean	std	mean	std	mean	std	mean	std
5	4.72	2.71	7.03	1.83	5.58	2.53	5.48	2.70
8	4.87	2.70	7.12	1.77	5.84	2.53	5.69	2.68
10	4.95	2.70	7.21	1.75	6.00	2.51	5.78	2.66
12	5.02	2.68	7.25	1.74	6.13	2.53	5.88	2.66
14	5.07	2.68	7.29	1.73	6.20	2.53	5.98	2.68
15	5.09	2.68	7.30	1.72	6.24	2.53	6.01	2.68
16	5.11	2.68	7.32	1.71	6.28	2.52	6.05	2.69
18	5.15	2.68	7.34	1.71	6.35	2.52	6.11	2.68
20	5.18	2.67	7.36	1.70	6.41	2.51	6.16	2.67
25	5.25	2.66	7.41	1.68	6.53	2.52	6.24	2.66

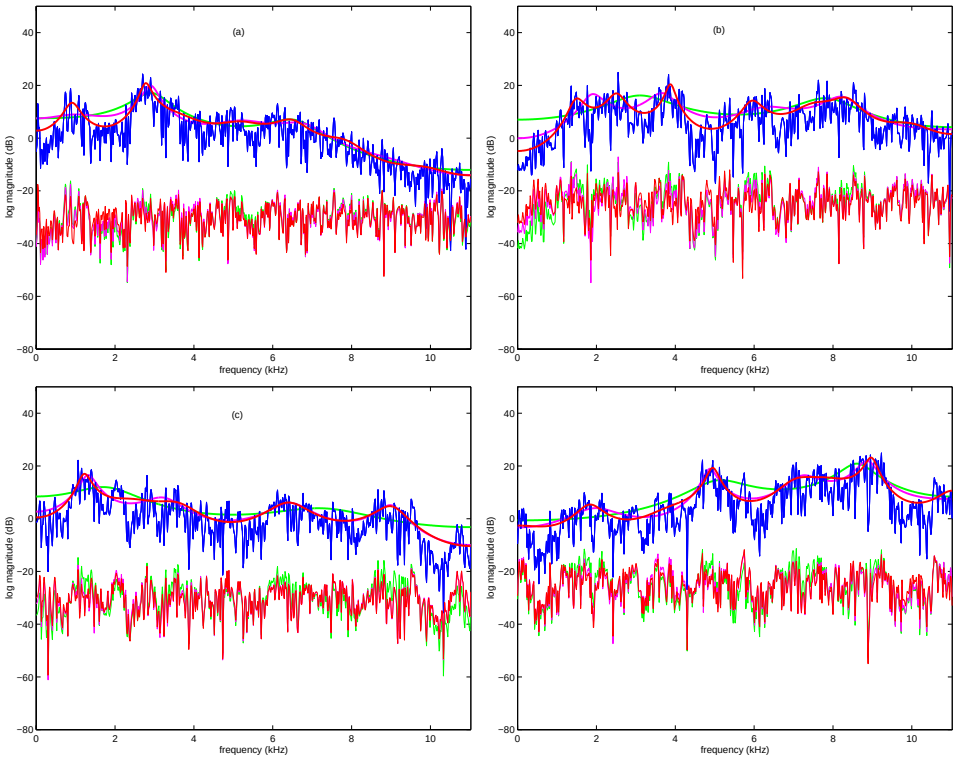

Fig. 1. Comparison of signal and error spectra from LP with variation of model order for a representative cough of (a) animal: chronic, (b) animal: chemical, (c) human: acute and (d) human: voluntary. The green, magenta and red fit corresponds respectively with $n_a = 5$, $n_a = 10$ and $n_a = 14$.

LP model orders and other pre-processing combinations, described in Subsection 3.1, didn't result in an over-all improved parameter-variance. Visual inspection of cough-waveform spectrograms however shows a very limited time-variation. So although the LP modelling is not optimal, the applied modelling might allow the localization of the harmonic resonances in the smoothed spectra. The effect of variation of the parameter estimates and the large standard deviations on the LP modelling performance of order $n_a = 14$ is illustrated in Figure 2 on (a) spectra and (b) poles of the same (animal: chronic) waveform-segment as used in plot (a) of Figure 1. The green, cyan, red and black results are obtained from respectively the parameters estimated on the exemplar sound, the class mean parameters as given in Table 2, the class mean parameters plus standard deviation and finally the class mean parameters minus standard deviation. From the presented spectra and poles it is easily to see the loss in model-accuracy obtained with the class mean parameters compared to the parameters estimated on the modelled sound, while modelling with the class mean parameters minus standard deviation even result in an unstable transfer function $\hat{H}(z)$ since one pole lies outside the unit circle. The models with parameters not estimated on the waveform-segment show large deviations for frequencies up to 4kHz. This paragraph illustrates the bad behavior of the retrieved mean LP parameters towards spectral features, which can be improved considering e.g. the Itakura-Saito distortion measure instead of the Euclidean distance.

Table 2. Mean LP parameter estimates a_k with associated relative standard deviation $\xi(a_k)$ (in %) for $n_a = 14$.

		a_1	a_2	a_3	a_4	a_5	a_6	a_7
animal	chemical	$-2.77\,10^{-1}$	$4.38\,10^{-1}$	$-1.27\,10^{-1}$	$4.21\,10^{-1}$	$1.04\,10^{-2}$	$2.72\,10^{-1}$	$3.11\,10^{-2}$
	$\xi(a_k)$	104.7	38.09	136.0	37.52	1788	55.40	469.4
	chronic	$-9.03\,10^{-1}$	$1.02\,10^{0}$	$-5.46\,10^{-1}$	$5.49\,10^{-1}$	$-2.72\,10^{-1}$	$2.27\,10^{-1}$	$-1.84\,10^{-1}$
	$\xi(a_k)$	18.81	15.40	37.70	38.24	60.55	70.32	64.31
human	voluntary	$-2.52\,10^{-1}$	$3.12\,10^{-1}$	$-2.66\,10^{-1}$	$3.51\,10^{-1}$	$-2.35\,10^{-1}$	$2.47\,10^{-1}$	$-6.45\,10^{-2}$
	$\xi(a_k)$	89.83	61.29	62.32	44.81	63.59	61.03	244.2
	acute	$-3.77\,10^{0}$	$5.34\,10^{-1}$	$-4.64\,10^{-1}$	$5.20\,10^{-1}$	$-3.55\,10^{-1}$	$3.72\,10^{-1}$	$-3.37\,10^{-1}$
	$\xi(a_k)$	76.97	34.12	57.58	40.06	55.03	56.32	54.80
		a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}
animal	chemical	$1.20\,10^{-1}$	$-9.49\,10^{-2}$	$5.33\,10^{-2}$	$-3.41\,10^{-2}$	$5.43\,10^{-4}$	$-1.71\,10^{-2}$	$-7.01\,10^{-3}$
	$\xi(a_k)$	61.33	275.6	94.71	2736	122.6	300.6	181.7
	chronic	$1.20\,10^{-1}$	$-9.49\,10^{-2}$	$5.32\,10^{-2}$	$-3.41\,10^{-2}$	$5.43\,10^{-4}$	$-1.71\,10^{-2}$	$-7.01\,10^{-3}$
	$\xi(a_k)$	109.6	104.0	224.4	250.6	$1818\,10^{1}$	379.3	976.1
human	voluntary	$2.21\,10^{-1}$	$-8.89\,10^{-2}$	$1.59\,10^{-1}$	$-1.55\,10^{-1}$	$5.14\,10^{-2}$	$-6.79\,10^{-2}$	$4.50\,10^{-2}$
	$\xi(a_k)$	64.35	178.0	85.36	84.62	238.3	175.9	177.6
	acute	$2.40\,10^{-1}$	$-1.88\,10^{-1}$	$2.20\,10^{-1}$	$-7.61\,10^{-2}$	$8.85\,10^{-2}$	$-9.24\,10^{-2}$	$1.54\,10^{-2}$
	$\xi(a_k)$	80.91	77.03	75.13	171.2	120.6	102.3	509.0

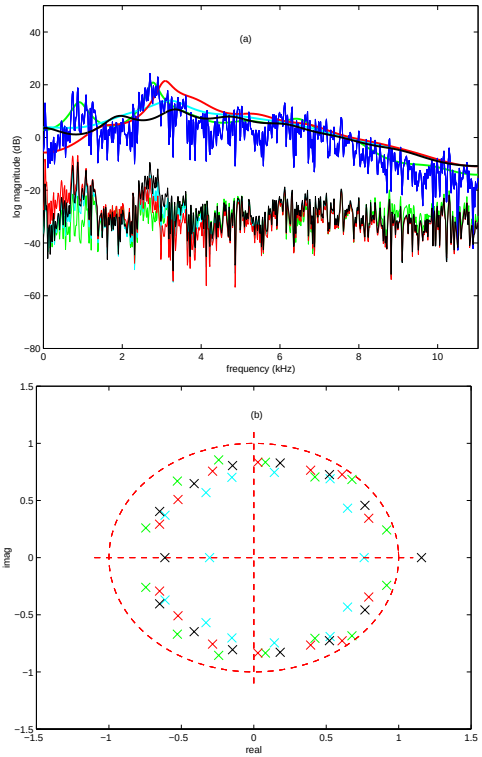

Fig. 2. Effect of variation of the parameter estimates for LP model of order $n_a = 14$ on spectra (a) and poles (b) for the animal chronic cough example also shown in part (a) of Figure 1. The green, cyan, red and black results are obtained from respectively the parameters estimated on the exemplar sound, the class mean parameters and finally the class mean parameters ± standard deviation.

3.4 *Formants from estimated LP parameters*

As pointed out in subsection 2.2 the positions and widths of the harmonic resonances can be estimated from the LP spectra using respectively Equations 4 and 5. Considering the loss in model-accuracy and consequently in peak-estimation of the LP-model with class mean parameters towards the parameters estimated directly on each waveform, as indicated in subsection 3.3, the formant frequencies, F_i, and bandwidths, B_i, are estimated directly on each waveform obtained with LP of $n_a = 14$. Table 3 returns the resulting inter-class mean formant frequencies, mean formant bandwidths and associated relative standard deviations ξ (in %). For both species 5 formants could be derived in a frequency-interval up to 10kHz or roughly one formant each 2kHz. The chosen model order of $n_a = 14$ can then be interpreted as 10 poles required to model the vocal tract leaving 4 poles to model the additional effects as e.g. glottal shaping and lip radiation. The large associated

Table 3. Mean formant frequency F_i, mean bandwidth B_i and associated relative standard deviations ξ (in %) for LP with $n_a = 14$.

	animal		human	
	chemical	chronic	voluntary	acute
F_1	1918	2130	1863	1789
$\xi(F_1)$	23	14	16	26
B_1	1073	1145	764	1177
$\xi(B_1)$	72	47	62	71
F_2	3480	3259	3560	3335
$\xi(F_2)$	14	6	12	10
B_2	1180	740	952	914
$\xi(B_2)$	71	57	63	73
F_3	4992	4593	5251	5129
$\xi(F_3)$	13	8	9	7
B_3	1314	1151	959	1171
$\xi(B_3)$	62	46	56	68
F_4	6609	6003	6832	6568
$\xi(F_4)$	7	6	7	5
B_4	1288	1291	1383	1185
$\xi(B_4)$	63	44	67	58
F_5	8113	7584	8151	7990
$\xi(F_5)$	5	5	4	4
B_5	1315	1589	1128	1169
$\xi(B_5)$	58	44	83	63

standard deviations on both mean F_i's and mean B_i's indicate a large variability on the mean values, which increases for higher frequencies. Except for the estimation of F_1, in general the mean formant F_i's estimated on animal waveforms are lowered compared to the frequencies obtained on human data. Except for the animal F_1-estimates, on both species the formants on healthy subjects are increased compared to the same formants for suffering subjects. Although Table 3 presents a quantitative estimation of the mean waveform formants and bandwidths for all 4 cough-classes discrimination of the 4 classes based on the pole-angles or formant-frequencies will be difficult due to the great relative standard deviations and therefore is not assessed here. However it could be remarked that the goal of class-classification might also be put forward as an other possible subjective criterion to determine the LP-model order.

4. CONCLUSION

Linear prediction acoustical modelling is assessed considering the prediction signal-to-noise ratio for distinct model orders and signal pre-processing steps on 4 distinct free field cough classes originating from animal and human species in different health conditions. For all cough waveform classes the model order is set to 14 and the best model performance is obtained after pre-emphasis of the waveform with a common first-order filter. For each cough class the vocal tract formants are estimated from the linear prediction parameters. Future research involves firstly the acoustic interpretation of the parameters and secondly the selection of a model-structure to decrease the large variance on the parameters.

REFERENCES

Irwin, S.I., M.J. Rosen and S.S. Braman (1977). Cough a comprehensive review, *Archives of internal medicine*, Vol.137, pp.1186-1991.

Korpas, J., J. Sadlonona and M. Vrabec (1996). nalysis of the cough sound: an overview, *Pulmonary Pharmacology*, Vol.9, pp.261-268.

Murata, A., Y. Taniguchi, Y. Hashimoto, Y. Kaneko, Y. Takasaki, and S. Kudoh (1996). Discrimination of productive and non-productive cough by sound analysis, *Internal Medicine*, Vol.37, pp.732-735.

Moreaux, B., D. Beerens, and P. Gustin (1999). Development of a cough induction test in pigs: effects of SR 48968 and enalapril, *J vet Pharmacol Therap*, Vol.22(6), pp.387-389.

Van Hirtum, A., and D. Berckmans (2000). Fuzzy approach for improved recognition of pig-coughing from continuous registration, Proc. 25th International ISMA Conference on Modal Analysis Noise and Vibration Engineering, Leuven, Belgium, pp.1535-1541.

Van Hirtum, A., and D. Berckmans (2001). Considering the influence of artificial environmental noise to study cough time-frequency features, Proc. 24th European Conference on Noise Control, Patras, Greece, pp.8.

Kobisch, M., and TNF. Friis (1996). Swine mycoplasmoses, *Revue scientifique et technique de l'office international des epizooties*, Vol.15(4), pp.1569-1605.

Rabiner L. and B.H. Juang (1993). *Fundamentals of speech recognition*. Prentice Hall, New Jersey.

Deller Jr., J. Proakis and J. Hansen (1993). *Discrete-time processing of speech signals*. Macmillan Publishing Company, Englewoods Cliff, N.J.