OBSERVATIONS OF A QUASI-PERIODIC INSTABILITY IN A LINEAR PREDIC--ETC(U)

AUG 77 D G NICHOL, R E BOGNER

WRE-TR-1879(W)
OBSERVATIONS OF A QUASI-PERIODIC INSTABILITY IN A LINEAR PREDICTION ANALYSIS OF VOICED SPEECH
A significant semiperiodic fluctuation of the vocal tract area functions derived by linear prediction of the speech waveform has been noted during apparently stationary voiced segments of speech. In one example some values of the area function varied over a range of 9:1 over a few pitch-periods. The phenomenon is attributed to 'beating' of the pitch-period and the time interval between successive computations which causes variations of the time relationship between glottal pulse and analysis window. This is supported by the fact that no fluctuations occur in the area function derived from natural or synthetic speech when the computation interval is equal to the pitch period. Any slight difference between the two leads to significant pulsations however. A simple theoretical model is used to show how the positioning of the analysis window can influence area function estimates.

The problem can be largely overcome by using longer time windows (greater than 2.5 pitch periods), or alternatively by averaging the area functions over several adjacent intervals.

Approved for Public Release.

* Department of Electrical Engineering, University of Adelaide.
<table>
<thead>
<tr>
<th>DOCUMENT CONTROL DATA SHEET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Document Numbers</strong></td>
</tr>
<tr>
<td>AR Number: AR-000-937</td>
</tr>
<tr>
<td>Report Number: WRE-TR-1879(W)</td>
</tr>
<tr>
<td><strong>Security Classification</strong></td>
</tr>
<tr>
<td>a. Complete Document: Unclassified</td>
</tr>
<tr>
<td>b. Title in Isolation: Unclassified</td>
</tr>
<tr>
<td>c. Summary in Isolation: Unclassified</td>
</tr>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td>OBSERVATIONS OF A QUASI-PERIODIC INSTABILITY IN A LINEAR PREDICTION ANALYSIS OF VOICED SPEECH</td>
</tr>
<tr>
<td><strong>Personal Author(s):</strong></td>
</tr>
<tr>
<td>D.G. Nichol</td>
</tr>
<tr>
<td>R.E. Bogner*</td>
</tr>
<tr>
<td><strong>Document Date:</strong></td>
</tr>
<tr>
<td>August 1977</td>
</tr>
<tr>
<td><strong>Total Number of Pages:</strong></td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td><strong>Number of References:</strong></td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td><strong>Corporate Author(s):</strong></td>
</tr>
<tr>
<td>Weapons Research Establishment</td>
</tr>
<tr>
<td><strong>Document (Wing) Series and Number</strong></td>
</tr>
<tr>
<td>Weapons Research and Development Wing TR-1879</td>
</tr>
<tr>
<td><strong>Imprint (Publishing establishment):</strong></td>
</tr>
<tr>
<td>Weapons Research Establishment</td>
</tr>
<tr>
<td><strong>Release Limitations (of the document):</strong></td>
</tr>
<tr>
<td>Approved for Public Release.</td>
</tr>
</tbody>
</table>

Security classification of this page: UNCLASSIFIED
A significant semi-periodic fluctuation of the vocal tract area functions derived by linear prediction of the speech waveform has been noted during apparently stationary voiced segments of speech. In one example some values of the area function varied over a range of 9:1 over a few pitch-periods. The phenomenon is attributed to "beating" of the pitch-period and the time interval between successive computations which causes variations of the time relationship between glottal pulse and analysis window. This is supported by the fact that no fluctuations occur in the area function derived from natural or synthetic speech when the computation interval is equal to the pitch period. Any slight difference between the two leads to significant pulsations however. A simple theoretical model is used to show how the positioning of the analysis window can influence area function estimates.

The problem can be largely overcome by using longer time windows (greater than 2.5 pitch periods), or alternatively by averaging the area functions over several adjacent intervals.
TABLE OF CONTENTS

1. INTRODUCTION
2. EXPERIMENTAL RESULTS
   2.1 The VTAF picture
   2.2 Observations
3. ANALYSIS OF WINDOW POSITION EFFECTS
4. EXPERIMENTAL DIAGNOSIS
5. SUPPRESSING THE FLUCTUATIONS
6. CONCLUDING REMARKS
REFERENCES

TABLE 1. POLE POSITIONS IN TERMS OF FORMANT FREQUENCIES AND BANDWIDTHS (AFTER FANT(REF.13))

LIST OF FIGURES

1. Vocal tract area function for sentence “Speak to me now, bad kangaroo!”
2(a) Speech waveform for /ae/
(b) Vocal tract area function for /ae/
3. Effect on windowed signal $s_w(n)$ of relative shift between signal and window
   (a) Window centred on signal
   (b) Window delayed with respect to signal
4. Approximation to a delayed window $w_d(n) = 0.58(1 + \cos \frac{2\pi}{128}(n-61))$ by a rising exponential $0.014 \times 1.26^n$
5(a) Physical tube model
(b) Signal flow model
6. Pitch period and computation interval are equal but with different relative positions in each column
7. Changing position of pitch pulse within window leads to pulsations when $n_c = n_w$
8. Increasing $n_w$ to 3.5 $n_p$ suppresses the fluctuations, for synthetic vowel /ae/
9. Vocal tract area function for “Speak to me now, bad kangaroo!” for window $n_w = 3.5$ times pitch period $n_p$
1. INTRODUCTION

Several attempts have been made recently to use linear prediction analysis of speech for isolated word (ref. 1, 2) and spoken digit recognition (ref. 3, 4). The feature chosen for the recognition algorithm in these studies was the set of linear prediction coefficients. It is well known that an estimate of the vocal tract area function can be derived from these coefficients (ref. 5, 6, 7) and the present paper arose from a study of the usefulness of this function for both speech and voice recognition. Because of the extensive information available from phonetic and articulatory studies of speech production it was believed that the vocal tract area function (VTAF) would be an advantageous feature for the pattern recognition process. To test this idea, and also to compare the various formulations of the linear prediction models, it was decided to display the VTAF as an intensity-modulated picture of vocal tract position versus time, with the area plotted as a grey-level. This is of course a similar display to the well known spectrogram. We shall refer to these displays as VTAF pictures.

Figure 1 shows a typical VTAF picture produced from real speech. The most obvious feature of this picture is the strong pulsations seen, for example, at the intervals labelled 2, 5 and 8. It is believed these pulsations are artifacts of the analysis as no evidence of the m is apparent in the time series or spectrogram.

Before discussing this phenomenon in more detail we describe briefly the production of the pictures.

2. EXPERIMENTAL RESULTS

2.1 The VTAF picture

The first linear prediction model used in this study was that due to Wakita (ref. 7). This is so-called auto-correlation technique and was chosen because, for non pitch-synchronous analysis, these formulations are generally more stable and robust than the "covariance" methods although for pitch-synchronous analysis the latter are capable of giving better estimates of the actual vocal tract (ref. 8, 9).

Suppose that the anti-aliasing filtered speech signal is sampled at frequency $f_s = 1/T$, and that $n$ samples are included in each autocorrelation window and that a new computation of the VTAF is made every $n$ samples.

If $m_1$ linear prediction coefficients are used then $m_V = m_1 + 1$ vocal tract areas are produced at time intervals of $t_c$ where

$$ t_c = n T \quad (1) $$

Denoting the array of vocal tract areas $a(t)$ obtained at time $t$ as a vector $\alpha(t)$ we have

$$ \alpha(t) = a_1(t), a_2(t), a_3(t), \ldots, a_{m_V}(t) \quad (2) $$

In $n$ successive estimates of $\alpha(t)$ are evaluated then the resulting sets of these $\alpha(t)$ may be regarded as an $(m \times n)$ matrix

$$ \tilde{\alpha} = \begin{bmatrix}
    a_{11} & a_{12} & a_{1n} \\
    a_{21} & a_{22} & a_{2n} \\
    \vdots & \vdots & \vdots \\
    a_{m1} & a_{mn} \\
\end{bmatrix} \quad (3) $$

where we have written $a_{mn}$ for $a_{m}(nT)$. 
This matrix can be plotted as an \((m \times n)\) digital picture where the grey levels are assigned from some mapping of the values of the elements of \(A\) to the set of grey levels.

To produce figure 1 the values \(f = 8192\) Hz, \(m = 9\), \(n = 64\) and \(n = 1024\) were used. Now a \(9 \times 1024\) picture is a very cumbersome shape and so this was split into eight \(9 \times 128\) subpictures, which for display purposes were interpolated (by a two-dimensional Fast Fourier Transform) into eight \(36 \times 512\) subpictures. These eight subpictures were plotted, one below the other, as in figure 1 on an intensity-modulated CRT. The bottom of each subpicture represents the glottis and the top the lips. The grey levels have been assigned such that the larger the area the greater the whiteness. Thus the point of maximum constriction is the darkest region in each column. It should be noted that the Wakita model assumes a constant glottis area and thus the lower edge of each picture is a constant grey level. Some regions of the picture are blank. This is due to use of an energy-detecting algorithm which assigns arbitrary zero levels to the VTAF's when the total signal occurring in the time series window is below a given threshold (as during silences between utterances). Each subpicture represents 1 s of real time and thus 8 s is shown overall.

2.2 Observations

In figure 1 the occurrence of periods of pulsations is easily observed; the most obvious of these are indicated in the picture. The utterance shown in this picture is the phrase “Speak to me now, bad kangaroo!” repeated three times by an Australian female speaker. The observed pulsations occur during constant vowel segments where little or no real change in the vocal tract is occurring.

This is supported by an examination of the speech time series corresponding to the utterance. Figure 2(a) shows the speech waveform corresponding to the /ae/ in “bad” which for this speaker is remarkably stationary. The corresponding VTAF’s are plotted in figure 2(b) and are clearly fluctuating, an effect which does not auger well for using the VTAF’s in any automatic speech recognition process. The plots in figure 2(b) are in fact the square-root of the VTAF and thus the estimates of area actually vary by the order of 9:1 during this apparently stationary segment.

We had not observed this phenomenon previously, even though several VTAF pictures of Australian male speakers repeating the same phrase had been made. This suggested that the phenomenon may be sensitive to pitch period.

Now for linear prediction models of the Wakita type the analysis begins by windowing the time series by a Hanning weighting function. The only parameter which is changing during a stationary segment is thus the position of the glottal pulse within the Hanning window (unless the analysis is pitch-synchronous). We shall now examine this effect and show it can cause the observed phenomenon.

3. ANALYSIS OF WINDOW POSITION EFFECTS

We examine the effect of the time relationship between the autocorrelation window and the speech waveform (figure 3). To facilitate analysis we use a model comprising a second order (two junction, three section) vocal tract yielding an impulse response of the form

\[
h(n) = r^n \cos n\omega T, \quad n = 0, 1, 2, \ldots
g(4)
\]

We use a window of the form

\[
w(n) = \frac{1}{2} + \frac{1}{2} \cos \frac{2\pi(n-d)}{N}, \quad n = \frac{N}{2} \text{ to } \frac{N}{2}
g(5)
\]

where \(d\) is the delay whose effect is of interest, and \(N\) the duration of the window is large compared with \(1/(r-1)\), i.e. the interval over which the impulse response has significant magnitude. The excitation is taken to be a unit pulse, and thus the model speech signal \(s(n)\) is the same as \(h(n)\).

This model is not realistic, but it does aid our appreciation of effects which can arise, and yields a sufficient explanation of our observations.

Figure 3 shows effects on the windowed signal \(s(n)\) variation of the delay. Two cases are shown, viz.
(a) Centred window, in which \( w(n) \approx 1 \) over the effective duration of \( s(n) \), i.e. we have \( s_w(n) \approx s(n) \).

(b) Delayed window, in which the curved rise of \( w(n) \) progressively magnifies the signal, producing a compensation of the damping of \( s(n) \).

Figure 4 shows for example the shape of the window

\[
w_d(n) = 0.5 \left[ 1 + \cos \frac{2\pi}{128}(n-61) \right]
\]

compared with the exponential \( Rr_1^n \) with \( R = 0.014 \) and \( r_1 = 1.26 \), which were chosen in an ad hoc manner simply for demonstration purposes. The exponential appears to be a reasonable approximation to \( w_d(n) \).

We see that for \( s(n) \) of the form

\[
s(n) = r^n \cos n\omega T
\]

with \( r < 1 \) the delayed window would cause \( s_w(n) \) to be approximately

\[
s_w(n) = s(n) w_d(n) \approx R (r_1 r_s)^n \cos n\omega T
\]

Now, for speech sampled at 10 000 s\(^{-1}\), the value of \( r_s \) is likely to be in the range 0.985 to 0.9 (i.e. approximately 50 Hz to 300 Hz formant bandwidth respectively). Thus, the apparent value of the ratio corresponding to \( r_s \) in \( s(n) \) as given by (7) becomes the value \( r_1 r_2 \) in (8) and the latter may be grossly in error, and even exceed unity as with the example values of \( r_1 = 1.26 \) and \( r_s = 0.9 \).

Next we study the effect of a discrepancy in the value of \( r_s \) on the area function of a model vocal tract.

Figure 5 shows an acoustic tube (or transmission line model of a vocal tract in which there are three sections of area \( a_m \), \( m = 0 \) at the lips end, 1 for the middle section and 2 at the glottis. There are thus two junctions whose volume velocity reflection coefficients \( \mu_m \), \( m = 1 \) and 2 are given by (ref.10)

\[
\mu_m = \frac{a_m^{-1} \cdot a_m}{a_m^{-1} + a_m}
\]

The termination at the lips is assumed to be equivalent to a tube section of infinite area, resulting in a volume velocity reflection coefficient of -1. For convenience in analysis we associate all the delay (i.e. sum of delays for forward and backward travelling waves) with the backward travelling wave in each section. The physical model of figure 5(a) may then be represented by the signal flow model of figure 5(b).

Analysis of this model shows that the transfer function \( H(z) = \frac{U_L(z)}{U_G(z)} \) is given by

\[
H(z) = \frac{(1+\mu_1)(1+\mu_2)}{1+z^{-1} \mu_1 (1+\mu_2) z^{-2} \mu_2}
\]

\[
= \frac{a_0 a_1}{(a_0 + a_1) (a_1 + a_2)} \cdot \frac{1}{1+z^{-1} \left( \frac{2a_1}{a_1+a_2} \frac{a_0 \cdot a_1}{a_0 + a_1} + z^{-2} \frac{a_2 \cdot a_1}{a_2 + a_1} \right)}
\]
The impulse response of this system is of the form

\[ h(n) = h(0) r^n \cos n\omega T \]

(12)

where

\[ h(0) = (1 + \mu_1)(1 + \mu_2) = \frac{4a_0 a_1}{(a_0 + a_1)(a_1 + a_2)} \]

(13)

\[ r^2 = \mu_2 = \frac{a_2 - a_1}{a_2 + a_1} \]

(14)

and

\[ \cos \omega T = \frac{1}{2r} \mu_1 (1 + \mu_2) = \frac{1}{2r} \left( \frac{2a_2}{a_1 + a_2} \right) \frac{a_0 - a_1}{a_0 + a_1} \]

(15)

From (14) we see that, for this two junction model, the damping ratio \( r \) depends only on the reflection coefficient of the junction closest to the glottis, i.e. on the area ratio at this junction. We might query the physical meaning of the possibility that \( a_2 - a_1 < 0 \) i.e., \( r^2 < 0 \) in (14). Detailed analysis shows that the impulse response is then not oscillatory, corresponds to real poles, and is not of interest in the present study.

To apply this two junction model to realistic speech parameters we set the length of each section equal to half the length of the vocal tract, i.e. about 9 cm. The tie in with the previous discussion of the 10 000 Hz sampling rate, it is convenient to let each of the sections be equivalent to an each way delay of \( 3 \times 10^{-2} \) s. The delay \( T \) in the second order model described by equation (8) is thus \( 6 \times 10^{-4} \) s, and the relevant values of \( r \) for use in these equations are \( (r \_r \_1)^6 \) or \( (r \_r \_6)^6 \). Of course this change in fact replaces the second order system by a 12th order system if the original sampling rate is maintained, since denominator factors of \( H(z) \) in (12) of the form

\[ (1 - z^{-1}re^{j\omega T}) \]

are replaced by factors of the form

\[ (1 - z^{-6}e^{6j\omega T}). \]

Each of these factors results in 6 poles, but the base pole of each is the same as previously, i.e. at \( z = re^{j\omega T} \).

From (14) we find

\[ a_2 = \frac{1 + r^2}{1 - r^2} \]

(16)

and

\[ \frac{da_3}{a_2} \bigg|_{a_1} = \frac{4r}{1 - r^2} \]

(17)

and

\[ \frac{da_3}{a_2} \bigg|_{a_1} = \frac{4r}{1 - r^2} \]

(18)
From (17) and also (18) we see that the proportional variation of either area $a_1$ or $a_2$ with $r$, while the other is fixed becomes very great as $r \rightarrow 1$. We found earlier that the effect of the delayed window on the apparent damping was sufficient to make $r$ pass through unity, and thus the system may incur such great sensitivities. For example, varying $r$ from 0.90 to 0.99 causes $\frac{a_2}{a_1}$ to change from 9.53 to 99.5. Clearly this effect is sufficient to account for variations as large as those observed in Section 2.2.

For completeness, we study the effect at the first junction. We note that $\omega T$ is not affected by the window delay phenomenon, and thus we set $d(\cos \omega T) = 0$ when differentiating (15).

We find

$$a_0 = a_1 \frac{r}{1 + r^2} \cos \omega T$$

(19)

and

$$\frac{da_0}{a_0} \left|_{r=1} \right. = \frac{1 - r^2}{r(1 + r^2)}$$

(20)

which shows that the area ratios at the first junction are not strongly influenced by $r$.

Note also that there is a gross effect on the initial value $h(0)$ or the windowed response. Via (13) we see that this can affect the product of the junction transmission coefficients i.e. $(1 + \mu_1)(1 + \mu_2)$. The effect on a particular feature however is not explicit.

For more complex vocal tract models, the effects are more complex, but we have demonstrated a sufficient mechanism to account for the observations. One previous study(ref.11) showed that under moderate variation of the pole dampings in a 5 pole signal, the resultant VTAF retained its gross features, but underwent a gradual smooth change.

4. EXPERIMENTAL DIAGNOSIS

To test these ideas, synthetic vowels were generated (in the computer) using an all-pole filter and known excitation function. Details of the synthesis algorithm used are given in Rogers(ref.12). The four poles used were derived from the values of formant positions and bandwidths given by Fant(ref.13) (Table 1).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/ae/</td>
<td>616</td>
<td>57</td>
<td>1072</td>
<td>72</td>
<td>2430</td>
<td>130</td>
<td>3410</td>
<td>175</td>
</tr>
<tr>
<td>/e/</td>
<td>432</td>
<td>39</td>
<td>1959</td>
<td>95</td>
<td>2722</td>
<td>170</td>
<td>3500</td>
<td>325</td>
</tr>
<tr>
<td>/i/</td>
<td>222</td>
<td>60</td>
<td>2244</td>
<td>75</td>
<td>3140</td>
<td>240</td>
<td>3700</td>
<td>230</td>
</tr>
<tr>
<td>/p/</td>
<td>510</td>
<td>54</td>
<td>900</td>
<td>65</td>
<td>2400</td>
<td>100</td>
<td>3220</td>
<td>135</td>
</tr>
<tr>
<td>/N/</td>
<td>231</td>
<td>69</td>
<td>615</td>
<td>50</td>
<td>2375</td>
<td>110</td>
<td>3320</td>
<td>115</td>
</tr>
</tbody>
</table>
Figure 6 shows plots of the VTAF's obtained for the synthetic vowel /æ/ when the pitch period $n_p$ has been made equal to the computation interval $n_c$. In each of the four columns however the 'phase' of the excitation function relative to the computation window is different (as indicated in the figure). Clearly the areas calculated vary with this phase. This means that when $n_p \neq n_c$ the area calculated from a constant waveform will fluctuate as the position of the excitation impulse changes within the window. Figure 7 shows this happening when $n_p = 0.8 n_c$ for five different synthetic vowels.

The reason this effect had not been observed in previous VTAF pictures of male speakers is believed to be that the computation interval used (64 samples, equivalent to 7.81 mS at 8192 Hz sampling rate) is quite close to the pitch period of the speakers analysed. Thus, fluctuations are not observed as the excitation function remains in a nearly constant position in the Hanning window. It should be remembered however that the errors may still be present in the analysis but not show up as fluctuations. For the female speaker $n_p \approx 0.9 n_c$ and the fluctuations are obvious (figure 1). This interpretation is supported by the fact that fluctuations did appear in male VTAF's that have been reprocessed with larger values of $n_c$.

5. SUPPRESSING THE FLUCTUATIONS

It appears from the above discussion that a partial cure for the problem of fluctuations would be to increase the size of the Hanning window used to estimate the autocorrelation function. This should improve the estimate of $r_s$. Figure 8 shows the synthetic vowel /æ/ as shown in figure 6 but with $n_w = 3.5 n_p$. We see that the variations are suppressed. To test this on real speech the VTAF picture (figure 1) was reprocessed with $n_w = 192$ (approximately $3.5 n_p$) and the result is shown in figure 9. The fluctuations have indeed been largely suppressed.

6. CONCLUDING REMARKS

The autocorrelation methods of linear prediction have a certain attraction in terms of robustness and economy of computing effort. We have shown that care must be taken in choosing lengths for the analysis, but provided this is done then consistent estimates of the vocal tract area are obtained. If Hanning windows of length $> 2.5 n_p$ are used, the resultant area functions appear to have the robustness desirable for automatic speech recognition, or for use in visual displays for speech training and phonetic studies.
<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Itakura, F.</td>
<td>“Minimum prediction residual principle applied to speech recognition”</td>
</tr>
<tr>
<td>2</td>
<td>Fujisaki, H. and Sato, Y.</td>
<td>“Evaluation and comparison of features in speech recognition”</td>
</tr>
<tr>
<td>3</td>
<td>Sambur, M.R. and Rabiner, L.R.</td>
<td>“A speaker-independent digital recognition system”</td>
</tr>
<tr>
<td>4</td>
<td>Rabiner, L.R. and Sambur, M.R.</td>
<td>“Some preliminary experiments in the recognition of connected digits”</td>
</tr>
<tr>
<td>5</td>
<td>Atal, B.S. and Hanauer, S.L.</td>
<td>“Speech analysis and synthesis by linear prediction of the speech wave”</td>
</tr>
<tr>
<td>6</td>
<td>Markel, J.D.</td>
<td>“Formant trajectory estimation from a linear least-squares inverse filter formulation”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCRL Monograph No. 7, Speech Communication Research Laboratory, Santa Barbara, 1971</td>
</tr>
<tr>
<td>7</td>
<td>Wakita, H.</td>
<td>“Estimation of the vocal tract shape by optimal inverse filtering and acoustic/articulatory conversion methods”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCRL Monograph No. 9, Speech Communication Research Laboratory, Santa Barbara, 1971</td>
</tr>
<tr>
<td>8</td>
<td>Chandra, S. and Lin, W.C.</td>
<td>“Experimental comparison between stationary and non-stationary formulations of linear prediction applied to voiced speech analysis”</td>
</tr>
<tr>
<td>9</td>
<td>Makhoul, J.</td>
<td>“Linear prediction: a tutorial review”</td>
</tr>
<tr>
<td>10</td>
<td>Markel, J.D. and Gray, A.H.</td>
<td>“Linear Prediction of Speech”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Springer-Verlag, Berlin 1976</td>
</tr>
<tr>
<td>11</td>
<td>Bogner, R.E. and Rogers, J.A.V.</td>
<td>“Determination of Vocal Tract Area Functions from a Pole Description of Speech Spectra”</td>
</tr>
<tr>
<td>12</td>
<td>Rogers, J.A.V.</td>
<td>“Determination of Articulatory Parameters from Speech Waveforms”</td>
</tr>
<tr>
<td>13</td>
<td>Fant, G.</td>
<td>“Acoustic theory of speech production”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mouton and Co., The Hague, Holland, 1960</td>
</tr>
</tbody>
</table>
Figure 1. Vocal tract area function for sentence “Speak to me now, bad kangaroo!”
Figure 2(a). Speech waveform for /ae/

Figure 2(b). Vocal tract area function for /ae/

Figure 3. Effect on windowed signal $s_w(n)$ of relative shift between signal and window
Figure 4. Approximation to a delayed window $w_d(n) = 0.58(1 + \cos \frac{2\pi}{128}(n-61))$ by a rising exponential $0.014 \times 1.26^n$

Figure 5(a). Physical tube model

Figure 5(b). Signal flow model
Figure 6. Pitch period and computation interval are equal but with different relative positions in each column.

Figure 7. Changing position of pitch pulse within window leads to pulsation when $n_c = n_w$. 
Figure 8. Increasing $n_w$ to $3.5 \frac{n}{p}$ suppresses the fluctuations, for synthetic vowel /æ/
Figure 9. Vocal tract area function for “Speak to me now, bad kangaroo!” for window $n_w = 3.5$ times pitch period $n_p$. 
DISTRIBUTION

EXTERNAL

In United Kingdom

Defence Scientific and Technical Representative, London

In United States

Counsellor, Defence Science, Washington

In Australia

Chief Defence Scientist

Executive Controller, Australian Defence Scientific Service

Superintendent, Central Studies Establishment

Controller, Policy and Programme Planning Division

Superintendent, Defence Science Administration Division

Chief Superintendent, Aeronautical Research Laboratories

Chief Superintendent, Materials Research Laboratories

Defence Library, Campbell Park

Library, Aeronautical Research Laboratories

Library, Materials Research Laboratories

Director, Joint Intelligence Organisation (DDSTI)

Superintendent, RAN Research Laboratory

Dr J.F. O'Callaghan, CSIRO Computing Research Division

Professor S. Kanefi, Australian National University, Canberra

Dr I.D.G. McLeod, Australian National University, Canberra

Dr B. Miller, Australian National University, Canberra

Professor B.D.O. Anderson, Newcastle University, N.S.W.

Professor F. Hirst, Adelaide University

Defence Information Services Branch (for microfilming)

Defence Information Services Branch for:

  United Kingdom, Ministry of Defence, Defence Research Information Centre (DRIC)

  United States, Department of Defense, Defense Documentation Center

  Canada, Department of National Defence, Defence Science Information Service

  New Zealand, Department of Defence

  Australian National Library

INTERNAL

Director

Chief Superintendent, Weapons Research and Development Wing

Superintendent, Electronics Division

Superintendent, Propulsion and Marine Physics Division

Superintendent, Systems Analysis Division
Senior Principal Research Scientist, Marine Studies Composite  43
Principal Officer, Marine Physics Group  44
Principal Officer, Underwater Detection Group  45
Dr D.V. Wyllie, Marine Physics Group  46
Dr A.K. Steele, Underwater Detection Group  47
Dr A. Werner, Underwater Detection Group  48
Dr S.H. Anderson, Cybernetic Electronics Group  49
Mr J.S. Bell, Marine Physics Group  50
Mr J.C. Andrews, Marine Physics Group  51
Dr D.J. Parker, Systems Integration Group  52
Authors  53 - 54
P.M.D. Library  55 - 56
W.R.E. Library  57 - 58
Spares  59 - 65