Proceedings of the Institute of Acoustics

VOCAL TRACT ACOUSTIC FUNCTIONS FOR VOICED SOUNDS

P.O.A.L. DAVIES

Institute of Sound and Vibration Research, University of Southampton, Southampton SO9 5NH

SUMMARY

This contribution describes the development and performance of realistic models for calculation of the acoustic transfer between glottis and lips during speech. Data input includes the relevant tract geometry including sinuses, with their surface properties, at some 30 or more stations along the tract. The resulting models have been successfully incorporated into articulatory synthesisers, where the calculated impulse response of the tract has been convolved with time varying models of the glottal source to synthesise speech.

INTRODUCTION

The vocal tract plays a fundamental role in the production of speech and other voiced sounds. Thus realistic and valid models of its acoustic characteristics play an essential role in the analysis, synthesis and recognition of such sound. Acoustic excitation is performed by modulation of the flow from the lungs either at the vocal folds and glottis or at other appropriately situated and controlled restrictions along the tract. The acoustic characteristics of the sound emitted at the lips and/or nose is further controlled by systematic and continuous adjustment to the shape of the whole tract by deliberate variation of muscle tension with associated movements of lips, jaws, tongue, velum and larynx. Thus, during articulation the vocal/nasal tract with its associated acoustic sources represents a time varying system, where any passive tract behaviour may be regarded as that of an acoustic filter. Although it is common practice to represent the acoustical behaviour of filters in terms of their spectral characteristics, this may not be appropriate in the present context. However, many time dependent systems may be analysed in discrete time steps at appropriate time intervals and also, so long as their relative time scales differ, be treated as a time varying or non-linear part coupled with a linear (or acoustic) part and the combination then represented as a hybrid system.

The vocal tract during voicing of vowels represents a familiar time varying hybrid system excited at the glottis. It can, for example, be modelled [1] as a combination of a relatively rapidly time varying part representing the dynamic behaviour at the glottis and a slowly time varying acoustic or linear part representing the vocal or vocal/nasal tract acting as an acoustic filter. The filter characteristics can be represented by a single or systematic time sequence of stationary acoustic transfer spectra between glottis and lips each corresponding to the successive geometric configurations of the tract. An inverse Fourier transform of each member provides a record of the changing impulse response of the filter at appropriate times, which can be convolved with the time varying glottal source characteristics. These include the influence of both unsteady mass flows and pressures. Similar procedures can be applied to other voiced and unvoiced elements of speech once the geometric configuration and corresponding source of excitation has been established. The result is then processed to synthesise speech with a clarity that depends in part on the effectiveness with which the two parts, source and filter, represent the associated fluid and boundary motions with their acoustic behaviour.

VOCAL TRACT ACOUSTIC FUNCTIONS FOR VOICED SOUNDS

An extensive sequence of filter models representing the acoustic characteristics of the vocal tract already exist in the literature. They are commonly described in terms of the electro-acoustic analogy and are thus restricted to one dimensional classical acoustic waveguide theory, neglecting mean flow and local cross-section shape, with other physical factors that can be significant. Full aeroacoustic modelling represents an alternative approach that includes all the physically significant factors [2], where three dimensional wave propagation is also included when relevant. A further alternative is provided by finite or boundary element models [3] though these neglect the aeroacoustics factors. The question which, if any, provides the most useful or realistic approach to acoustic modelling forms the subject of this contribution.

1. VOCAL TRACT GEOMETRY

The configuration of the vocal tract with the nomenclature of its elements is illustrated by the mid-sagittal section in Figure 1. Note too that the width and cross-section shape also vary with position. The velum is shown closed to correspond to the condition normally observed when vowels are produced. The boundary surfaces of the tract consist mainly either of muscular or soft tissues whose state can change from tense to lax and which also respond to acoustic excitation. The remaining surfaces are either hard (teeth) or closely backed by bone. Measurements are available in the literature presented as mid-sagittal sections with axial area distributions (area functions) and on occasion some useful cross sections, including sets of data derived during the sustained articulation of vowels. While such observations were undertaken, the subject attempted to repeat and maintain a consistent utterance. When experimental conditions allowed, though this seems unusual, simultaneous records were obtained of the acoustic characteristics of the emitted sound, otherwise precautions were taken to ensure that the corresponding acoustic records remained representative.

Measurements of the vocal tract geometric configuration by Fant [4] and by Chiba and Kajiyama [5] derived from X-ray photographs, supplemented by other observations, provide most of the data considered here. Similar measurements [6] derived more recently from magnetic resonance imaging [MRI] have also been considered. In its simplest form [4, 5], the measured geometric data was summarised as a one dimensional area function, describing the axial variation in tract cross-section area along the central axis of the tract from lips to glottis. Area functions for the pyriform sinuses were included [5, 6] with some of the data. However, such functions ignore the curvature of both the axis and boundaries with the wide variations in cross-section shape that exist along them. On the other hand, a complete three-dimensional representation during articulation inevitably involves so many measurements that some distortions due, for example, to inconsistencies in utterance seems inevitable. The fact that both the curvature and presumably the length and shape distribution vary with the inclination of the head introduces further uncertainty in establishing a consistent and representative three dimensional description.

Furthermore, it is well established that the vocal tract varies widely, both in shape and overall scale, between individuals of the same sex, between the sexes and between adults and children. Also there are individual differences in the articulatory configuration associated with the same phonetic element of speech (phoneme). Despite this, there seem to be clear similarities in the area functions observed for three male individuals voicing the vowel /a/, as in Father, that are plotted in Figure 2. The linguistic backgrounds of the subjects were rather different, since one was Russian [4], one Japanese [5] and one American [6]. The fact that the first two were derived from X-ray photographs while the third with the MRI technique might explain the obvious difference in the corresponding area function

near the teeth, since such techniques are ineffective for teeth or thinly covered bone. The sharp discontinuities in the slope of the area functions shown in the figure obviously are spurious, since the actual tract boundary contours are continuous.

2. ACOUSTIC MODELLING

The acoustic modelling adopted for this study was based on well established and verified aeroacoustic methods [7] which include further appropriate modifications (i.e. reacting walls) and developments described in reference [2]. The normal set of assumptions adopted for classical acoustic theory are listed there, with the extent to which they were relaxed. For example the acoustic medium was treated as a viscous heat-conducting fluid with time averaged velocity u_0, rather than an ideal fluid at rest. Due account was included for the influence of flow separation and vorticity generation in relevant sections of the tract. Studies of the influence of the pronounced curvature of the axis on wave propagation at the frequencies of interest, indicated that it remained small and could normally be neglected.

The area functions obtained by MRI were evaluated at 5 mm intervals [6] along the tract axis and are represented by the continuous line in Figure (3). This provides the appropriate description of the duct geometry for a one dimensional acoustic model described later. Further MRI data on cross-section shape at various stations along the tract was used to develop a sequence of 3D elements with a somewhat simplified regular geometry to represent the tract, that are indicated by the dashed line in Figure 3. The influence of wall vibration and real gas properties on wave dispersion was evaluated with the corresponding hydraulic radius, $2S/P$, where S is the cross-section area and P the perimeter. The left and right pyriform sinuses are also shown marked L and R respectively. These regular elements were employed for an earlier acoustic prediction [2], which included appropriately contoured wave fronts, rather than the plane ones adopted for one dimensional acoustic calculations. The relevant details can all be found in reference [2], so are not repeated here.

3. ACOUSTIC PREDICTIONS

The frequencies of the first three formants corresponding to the vocal tract shown in Figure 3 were listed [6] as 595, 1006 and 2400 Hz respectively. These can be compared with predictions, with sinuses included, and with the 3 D modelling [2] of 590, 1010 and 2480 Hz respectively without flow and 600, 1012 and 2480 Hz respectively with a mean volume flow of 0.2 l/s, assumed as a representative value. The results demonstrate that the influence of flow is less than one percent, indicating that its presence may probably be neglected, at least as far as the frequencies associated with voicing are concerned for vowels where the tract remains relatively unrestricted, a result that seems realistic and hardly surprising. The agreement between observations and predictions seems similarly close, at least for the first two formants. Similar agreement between predictions and observations [2] was found for the vowel /ɜ/ where the description of tract geometry [4] also included a sequence of cross-sectional shapes.

However, descriptions of tract geometry often omit the cross-sectional shape. Then one is obliged to adopt a one dimensional acoustic model based simply on the area function data. Such models seem fairly common and consist of an appropriate sequence of straight tubes of constant area [4, 5] the number depending on the extent to which realism of representation was attempted. Though

remaining one dimensional, the acoustic modelling adopted here takes due account of the hydraulic radius together with the local rate of change of area along the tract, with a different type of modelling for high and low rates of change respectively. This was in addition to the inclusion of real gas and flow effects with that of reacting boundaries in the modelling. Acoustic predictions with the simple model using the sequence of elements shown by the solid curves in Figure (3) yielded formant frequencies of 560, 1075 and 2525 Hz respectively that now deviated more obviously from those observed for the subject, since discrepancy between observations and predictions had now increased to around six percent.

Similar calculations were performed on the vocal tract for the vowel /a/ with geometry described in reference [4] as illustrated in Figure (4). Since no cross-section shapes or details of the geometry of the sinuses was provided, one dimensional acoustic modelling was adopted yielding the first three predicted formant frequencies at 680, 1130 and 2490 Hz respectively compared with the observed values [4] of 700, 1080 and 2600 Hz respectively. The predicted frequencies of the first and third formants were raised slightly when sinuses with an "estimated" geometry were included.

Finally, one dimensional acoustic modelling was applied to the measurements for the same vowel given in reference [5]. A sequence of cross-section shapes for the tract with area functions for the sinuses were both included in the data. Though a line spectrum of the corresponding sound produced by the subject was recorded, the data had not been processed to define the formant frequencies. One can however compare the predicted formant frequencies from the one dimensional model with those obtained with finite element and boundary element models presented in reference [3]. Some of the results of existing observations with predictions that are currently available for /a/ are summarised in table I, which is still incomplete.

The first set of comparisons include the result of digital and electro-acoustic analogue modelling respectively also without sinuses labelled BESK and LEA, described by Fant [4]. The predictions with the new one dimensional model seem to be slightly more realistic. The third set [6] have already been discussed, but one might recall that the approximate three dimensional modelling gave more realistic predictions. The second set of comparisons in the central rows of the table are all based on the geometry described in reference [5], reproduced in reference [3], accompanied by acoustic predictions omitting the pyriform sinuses. The three dimensional acoustic predictions calculated with finite element [FEM] and boundary element [BEM] methods and reacting walls [3] were extracted from among the results listed there. The one dimensional acoustic predictions are the results of new calculations for the same geometry [5] including mean flow but either omitting or including sinuses. Comparison of the four sets of results give little indication of a significant trend, except for the third formants. This suggests that the influence of tract curvature on wave propagation can probably be neglected, at least at the lower formant frequencies. Any significant influence of tract curvature on the predicted formant frequencies for this vowel will be illustrated more clearly after that section of the table is completed, firstly by calculating the values of the corresponding observed formant frequencies and secondly by performing acoustic predictions in three dimensions with straight ducts.

This paper represents an intermediate stage in the investigation since clearly much remains to be done. Similar calculations including all the measured data available in the references cited for other vowels is an obvious next step. Extension to other types of voiced sounds is another. The one dimensional model also provides useful facilities for studying the relation between formant frequency and the tract's geometrical features by introducing small but appropriate systematic adjustments to

the area function, and/or sinuses. Such a study should be illuminating. All one can say at present is that the results remain encouraging.

4. ACKNOWLEDGEMENTS

The author is grateful to the Leverhulme Trust for their encouragement by the award of an Emeritus Fellowship and to Dr C H Shadle for helpful discussions and advice.

5. REFERENCES

[1] M M SONDHII and J SCHROETER "A hybrid time-frequency domain articulatory speech synthesiser". I.E.E.E. Trans. ASSP 35, 955-967 (1987).

[2] I R TITZE (ed) "Vocal fold physiology, frontiers in basic science". Chapter 3, P O A L DAVIES, R S MCGOWAN and C H SHADLE "Practical flow duct acoustics applied to the vocal tract" 93-142. Singular Publishing, San Diego (1993).

[3] T KAGAWA and R SHIMOYAMA "Boundary element models of the vocal tract and radiation field and their response characteristics". J. Sound Vib. 157, 385-403 (1992).

[4] G FANT "Acoustic theory of speech production". Mouton and Co., The Hague (1960).

[5] T CHIBA and M KAJIYAMA "The vowel its nature and structure. The vowel - its nature and structure part 3". The Phonetic Society of Japan (1958).

[6] T BAER, J C GORE, L C GRACCO and P W NYE "Analysis of vocal tract shape and dimensions using magnetic resonance imaging". JASA 90, 779-828 (1991).

[7] P O A L DAVIES "Practical flow duct acoustics". J. Sound Vib. 124, 91-115 (1988).

TABLE I, Formant frequencies, Hz, observed and predicted for /a/

	Observations		Acoustic predictions					
			three dimensional			one dimensional		
			[2]	[3] FEM	[3] BEM	No sinus	sinus	BESK LEA
[4]	*F*1	700				680	616	630
	*F*2	1080				1130	1072	1072
	*F*3	2600				2490	2470	2400
[5]	*F*1			710	720	725	705	
	*F*2			1240	1248	1275	1205	
	*F*3			2565	2579	2895	2855	
[6]	*F*1	595	600				560	
T.B.	*F*2	1006	1012				1075	
	*F*3	2460	2480				2525	

note the numbers in square brackets, all refer to the list of references.

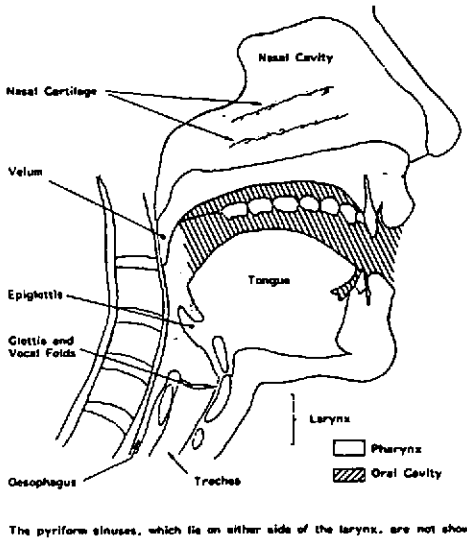

Figure 1 The vocal tract configuration and nomenclature

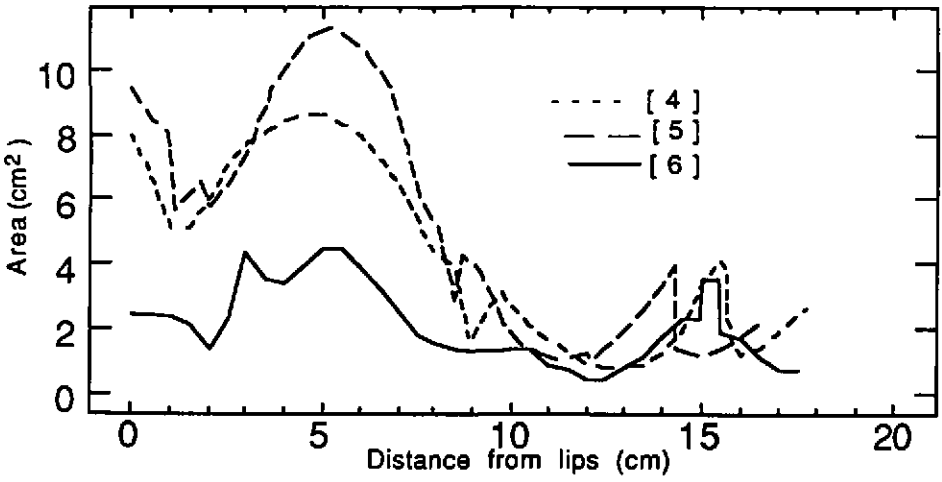

Figure 2 Area functions for the vowel /a/
[4] Russian, [5] Japanese, [6] American

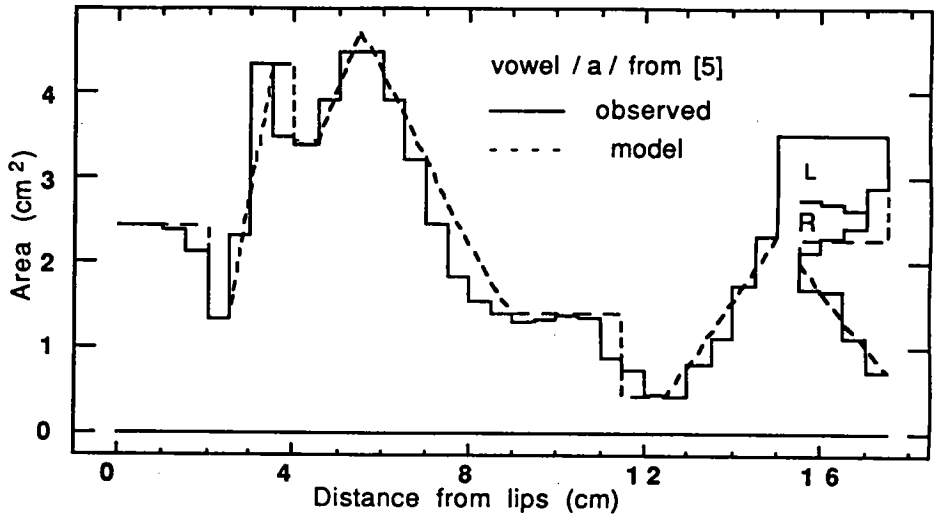

Figure 3 Area function (3D) adopted for acoustics

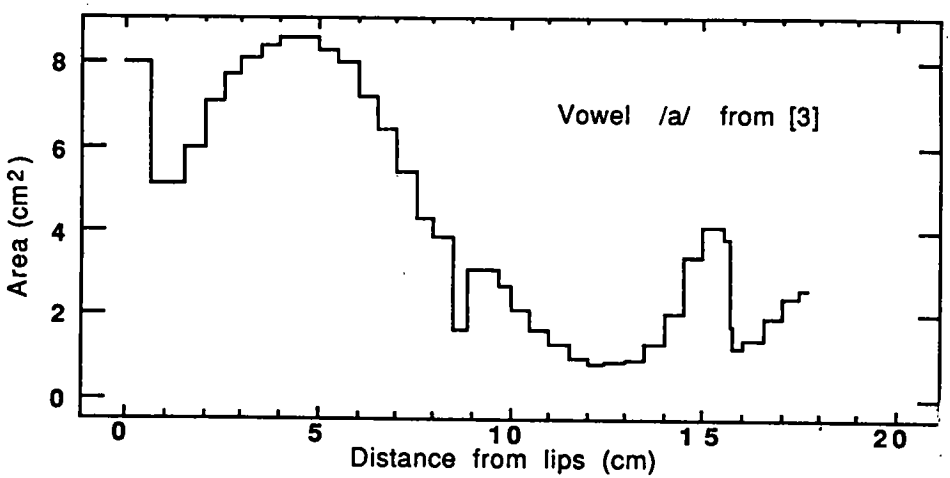

Figure 4 Area function (simple) adopted for acoustics