Between Formant Space and Articulation Space

L.J. Bonder
Amsterdam, the Netherlands

1. Introduction

One of the problems we meet in the study of speech production models is how we can describe the relation between vocal tract and generated sound. More specifically, for vowellike sounds, the question is: given a vocal tract shape, what are the formant values, and conversely, given a certain set of formant values, which shapes can produce these formants. The latter problem of determining shapes from formants will be referred to as the inverse problem.

In this paper we will sketch the relation between formants and shape of a model of the vowel tract. Finally, we will give an outlook on further research on this issue.

2. The n-tube model

The starting point in the description of the relation between formants and articulation is the modelling of the vocal tract as a lossless n-tube, i.e. a concatenation of n cylindrical tubes of equal length l, but different cross-sectional area (Dunn, 1950). The cross-sectional area of segment i is denoted by S_i (i=1,....,n). Further, we define the so-called k-parameters k_i, by putting $k_i = Si/S_{i+1}$ (i=1,...,n-1).
The propagation of sound through such a tube is described mathematically by the one-dimensional wave equation. The pressure and the volume velocity are considered continuous at the junctions of the segments. If we put together the n-tube model, the one-dimensional wave equation, and the continuity conditions, we get the so-called n-tube formula, a closed form expression relating shape to formants, and vice versa (Bonder, 1983a).

3. Properties of the n-tube model

As we have seen, the model has three important features: continuity of pressure and volume velocity, one-dimensionality, and its lossless nature. Beside these features there are some interesting consequences of the model (cf. Bonder, 1983a).

First, there is the modelling of the vocal tract in a non-continuous way as an n-tube. The consequence of this non-continuity is that, from an acoustic

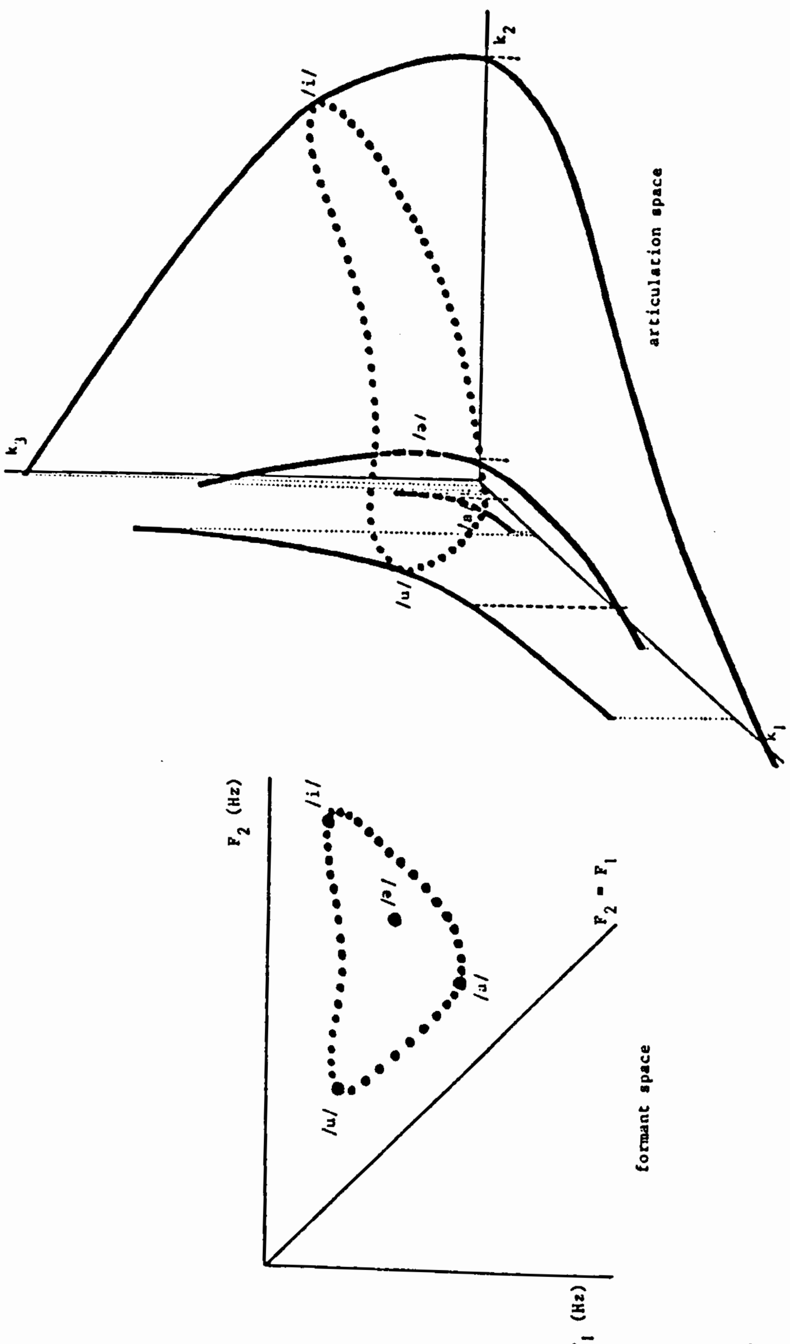

Figure 1. An n-tube model of the vocal tract. The k-parameters are $k_i=S_i/S_{i+1}$ $(i=1, \ldots, n-1)$.

point of view, only the first [½n]* formants of an n-tube can be taken seriously; the pattern of the higher formants is merely a repetition of the lower [½n] formants.

Secondly, the formant frequencies of an n-tube will not change if the n cross-sectional areas S_i are multiplied by the same factor. So one of the S_i can be taken as a reference for the other S_i. This means that an n-tube can be fully described by the n-1 parameters $k_i=S_i/S_{i+1}$ $(i=1,\ldots, n-1)$. The description of an n-tube in terms of its n-1 parameters k_i enables us to view such a tube as a point in the (n-1)-dimensional space spanned by these parameters. An n-tube P with k-parameters $k_1, \ldots, k_{n-1}$ is denoted by $P=(k_1 \ldots, k_{n-1})$. The space spanned by the k-parameters will be called 'articulation space'. In Fig. 2 the location of the straight 4-tube in the 3-dimensional space of k_1, k_2, k_3 is shown. As all S_i have the same value, it follows that $k_1=k_2=k_3=1$. So, the straight 4-tube is denoted by (1,1,1).

Thirdly, when calculating n-tube shapes from formant frequencies there are [½(n-1)] degrees of freedom. This means that we can choose freely the values of [½(n-1)] parameters k_i. For example, 4-tubes have one degree of

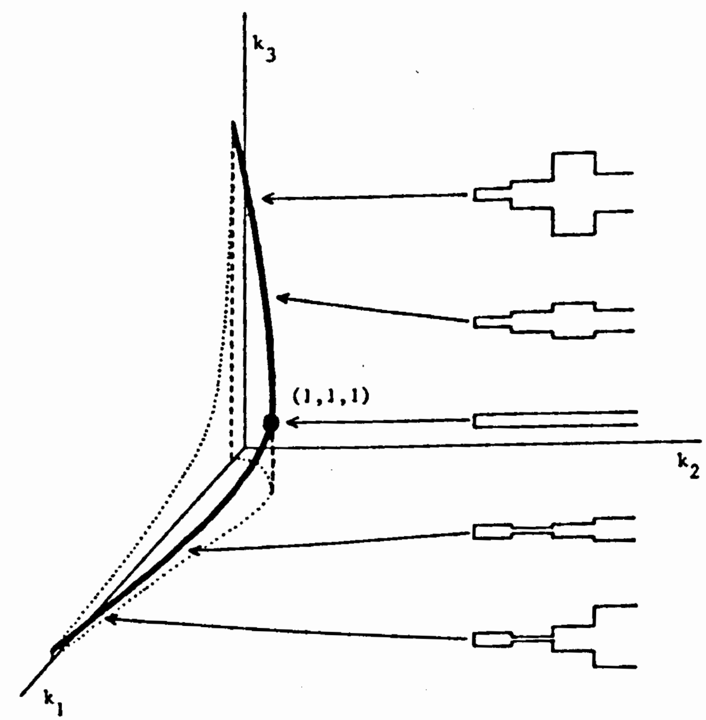

Figure 2. The 3-dimensional articulation space of 4-tubes spanned by the parameters k_1, k_2, k_3, and the location of the straight 4-tube (1,1,1) in it.

* [½n] is defined as the greatest integer number not greater than ½n.

freedom: $[\frac{1}{2}(n-1)]=1$. We observe that the inverse problem does not have a unique solution for n-tubes with more than two segments.

4. The Inverse Problem

The greatest contribution to the research on the inverse problem is from Atal, Chang, Mathews and Tukey (1978). They treated the subject numerically, and showed that there are many vocal tract shapes having the same formant frequencies. A disadvantage of their numerical inversion is that it does not show the structure of the relation between shape and corresponding formant pattern. We will briefly indicate when and how we can handle the inversion analytically (Bonder, 1983b).

The n-tube formula is the starting point for the attack on the inverse problem. The main step in our method of inversion is the decomposition of the n-tube formula, i.e. the replacement of the n-tube formula, which is an equation of degree n, by a set of $\frac{1}{2}n$ equations relating explicitly formants to shape of the tube. By means of these $\frac{1}{2}n$ equations we can solve the inverse problem analytically up to 10-tubes, but, the more segments, the more involved the calculus. For n-tubes consisting of more than 10 segments the inverse problem is no longer analytically solvable, in which case the problem has to solved numerically.

For 4-tubes the inversion is rather simple. The analytical inversion yields the following expressions from which the k-parameters can be determined if the formants F_1 and F_2 are known:

$$\begin{aligned} k_2 &= (-C_2k_1^2+C_1k_1-1)\,/\,((1+k_1)(1+C_2k_1)) \\ k_3 &= 1\,/\,C_2k_1 \end{aligned} \qquad (1)$$

where

$$\begin{aligned} C_1 &= \tan^2\tau F_1 + \tan^2\tau F_2 \\ C_2 &= \tan^2\tau F_1 \cdot \tan^2\tau F_2 \\ \tau &= 2\pi L/4c \end{aligned}$$

c being the velocity of sound, L (=4.1) the overall length of the 4-tube. It is obvious from equations (1) that we have one degree of freedom for 4-tubes: one parameter, k_1, has to be given a value in order to be able to compute the other two parameters k_2 and k_3. Expressions (1) can be used to calculate equivalent 4-tubes, i.e. tubes with the same formant frequencies, to a given length L. Each equivalence class consists of an infinite number of 4-tubes, all of them having the same length L. In the articulation space, an equivalence class turns out to be a continuous trace. In Fig. 3 we show the equivalence class of the straight 4-tube with formants F_1=500 Hz and F_2=1500 Hz. All equivalent tubes have the same overall length L=17.5 cm (which is about the

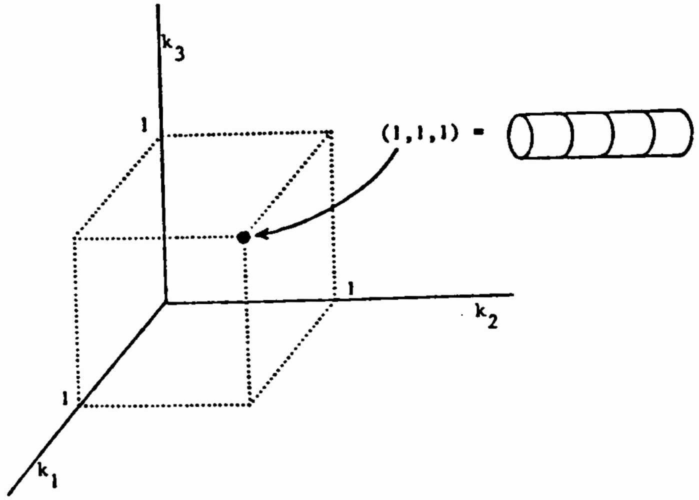

Figure 3. The equivalent class of the straight 4-tube (1,1,1) with F_1=500 Hz and F_2=1500 Hz and length L=17.5 cm (solid curve). At the right hand side the corresponding 4-tube shapes are shown.

average male vocal tract length). Of course,we cannot show the whole class but this finite subset gives us a good view of it.

5. Inversion Applied to the Vowel Triangle

In the way mentioned above we can calculate a trace in the articulation space for each point in the formant space. All traces in the articulation space look very much the same as the one in Fig. 3. In Fig. 4 we have sketched the traces in the articulation space corresponding to the three 'vertices' of the vowel triangle /u/, /i/, /a/. As we may see from Fig. 4, the structure of the vowel triangle is rather alike in both spaces. From this we conclude that our choice of the k-parameters as parameters of articulation seems to be adequate. The structure of the articulation space of Atal et al. (1978) is more complex. Besides, one of their dimensions is not contained in our space, namely the length L of the tube. From an acoustical point of view, the length L is not essential, as it is no more than a scaling factor in the formant space (for, if all the segment lengths are multiplied by the same factor α, the cross-sectional areas being unchanged, the overall length will change to αL and the corresponding formant frequencies F_i to F_i/α).

6. Outlook on Further Research

After this rough sketch of the relation between formant space and articulation space by means of the n-tube model of the vocal tract we might come to the question if there are preferential areas in the articulation space, and how

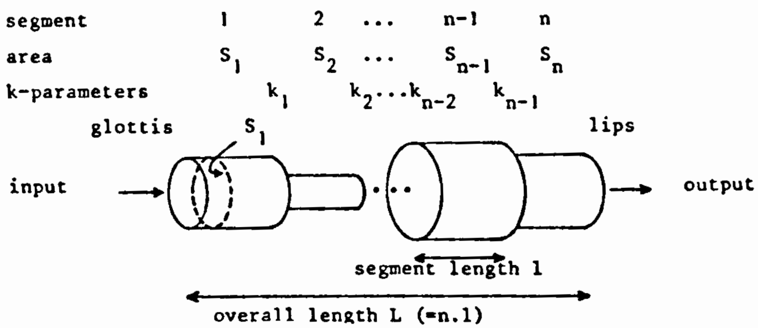

Figure 4. The vowel triangle in both formant space and articulation space. The solid curves in the articulation space are the equivalence classes of the vowels /u/, /i/, /a/, /ə/. In the articulation space we have drawn an arbitrary cross-section (solid dots) of the bundle formed by the equivalence class curves of vowel sounds. The curve corresponding to the /ə/ intersects this cross-section.

we can describe this phenomenon in terms of our k-parameters k_i. As a starting point in this direction we use the paper by Lindblom and Sundberg (1971). They suggested, on the basis of numerical experiments, that a principle of minimal articulatory antagonism between tongue and jaw might play an important role in the realization of isolated vowels. If we want to translate this mechanism into the language of our model we obviously have to define some measure with which we can quantatively indicate the resemblance of two n-tube shapes. As a measure of comparison between two n-tubes.

$$P=(k_1^{(p)},\ldots,k_{n=1}^{(p)}) \text{ and } Q=(k_1^{(q)},\ldots,k_{n-1}^{(q)})$$

we introduce in a forthcoming paper (Bonder: the MAD model)

$$d_{P,Q} = \left\{ \sum_{i=1}^{n-1} (k_i^{(p)} - k_i^{(q)})^2 \right\}^{1/2}$$

which is the euclidean distance between the two tubes P and Q in the (n-1)-dimensional articulation space.

The translation of the suggestion by Lindblom and Sundberg (1971) into our model seems to be that we have to look, in the articulation space, for the point on the trace corresponding to a vowellike sound that has minimum distance to the straight tube, the point with coordinates (1,1,1) in the 3-dimensional articulation space of k_1, k_2, k_3. This is the so-called MAD model, where MAD stands for Minimal Articulatory Difference.

Acknowledgements

This paper is based on research which was supported by the Foundation for Linguistic Research, funded by the Netherlands Organization for the Advancement of Pure Research Z.W.O. under Contract 17-21-08.

References

Atal, B.S. et al. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America* **63**, 1535-1555.

Bonder, L.J. (1983*a*). The n-tube formula and some of its consequences. *Acustica* **52**, 216-226.

Bonder, L.J. (1983*b*). Equivalency of lossless n-tubes. *Acustica* **52**, (to appear).

Dunn, H.K. (1950). The calculation of vowel resonances, and an electrical vocal tract. *Journal of the Acoustical Society of America* **22**, 740-753.

Lindblom, B.E.F. and Sundberg, J.E.F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America* **50**, 1166-1179.