10ème Congrès Français d'Acoustique
Lyon, 12-16 Avril 2010

The Gunnar Fant legacy in the study of vocal acoustics

Björn Lindblom[1], Johan Sundberg[2], Peter Branderud[1], Hassan Djamshidpey[1], Svante Granqvist[2]

[1]Phonetics Lab, Linguistics, Stockholm University, 10691 Stockholm, Sweden, lindblom@ling.su.se
[2]Dept of Speech Music and Hearing, Royal Institute of Technology, 10044, Stockholm, Sweden pjohan@speech.kth.se

Gunnar Fant made a number of seminal contributions to the acoustics of human speech and voice production. Early in his career – even before defending his dissertation and being appointed professor – he was already an established researcher of world renown. The main focus of his KTH laboratory was the acoustic analysis and synthesis of speech. The forte of this lab was the successful attempts to synthesize natural sounding speech electronically. The good results were due to the fact that Fant had used his broad understanding of circuits and acoustics to formulate the *source-filter theory*. This is an account that makes it possible to deduce the physical properties of speech from first principles. Today we know that it gives a highly accurate description of the speech signal and that it explains how vowels and consonants get their acoustic properties. It is general, language-independent and valid for both normal and disordered speech. It can be applied to the human voice in singing as well as in speech. In our presentation we will try to identify still outstanding questions that remain incompletely understood. As examples of such questions, issues arising in the quantitative modeling of the front cavity of the vocal tract will be addressed, notably the treatment of spread lips and the sublingual cavity formed under the tongue blade in coronal consonants..

1 Introduction

Gunnar Fant was one of the founding fathers of acoustic phonetics. His early work produced the source-filter theory which revitalized phonetics, the 'study of speech sounds'. In the course of time certain limitations have been recognized and addressed (14). Currently it creates the solid theoretical foundation on which the scientific study of speech and voice rests. In reflecting on his career (2, 3) Fant noted that the most quoted part of his 1960 monograph is the section on the ***nomograms*** relating formant frequencies to three parameters of a vocal tract tube model. This model compellingly demonstrates that the place and cross-sectional area of the main vocal tract constriction have greater acoustic relevance than the front-back and high-low dimensions of traditional vowel classification.

2 APEX – an articulatory model

As students of Fant we were greatly inspired by this work. In its spirit we focused on developing nomograms relating formant frequencies to articulator positions rather than to area-function parameters. This research resulted in the APEX model, a tool that we have found useful in studying the acoustic consequences of articulatory movements. Based on X-ray measurements, APEX is a system of rules that convert jaw opening, lip shape, larynx height, tongue blade (position and retraction/protrusion of apex) and tongue body (position and displacement from neutral) into formant frequencies. APEX uses independent control of the jaw and the tongue. Possible tongue shapes are specified relative to a neutral reference tongue. These choices offer a great deal of explanatory mileage (4, 5).

This project has also highlighted certain outstanding problems in the acoustic modeling of vocal tract shapes. For instance, it is clear that our understanding of vocal tract acoustics remains incomplete with respect to the treatment of ***raised tongue blade*** - which creates an acoustically relevant sublingual cavity - and ***spread lips*** - whose 3-D shape presents a special problem for 2-D area function models like APEX.

3 Where does the VT end?

3.1 Spread lips: Cylindrical (notched) tubes

As a rough approximation of "spread lips" we used cylindrical copper tubes whose inner diameters were 20, 25, 32 mm. Figure 1 illustrates these tubes. "Mouth corners" were simulated by cutting and filing the tubes at one end so as to produce wedge-like notches. The center columns of Figure 1 illustrate their geometry. The depth of the notches was varied between 0, 1, 2, 3, 4, and 5 cm. In the side view they produce straight lines; from the top or the bottom their contours produce a curved projection. Tubes without notches were 'long' and 'short'. Short tubes ended at the mouth corner; Long tubes extended from source to the top edge.

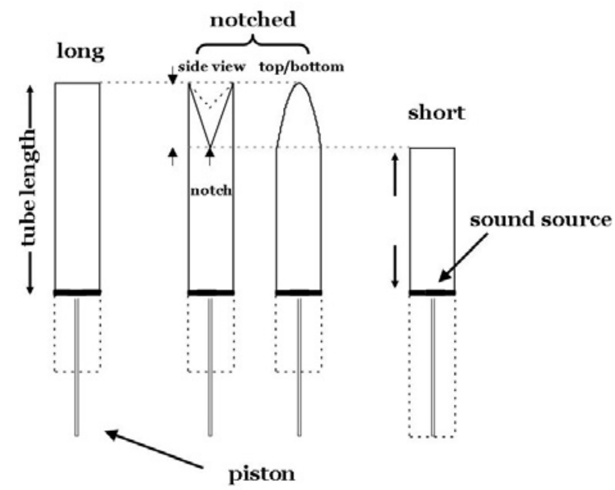

Figure 1: Cylindrical tubes for studying "spread lips".

The formant frequencies of these resonators were measured by means of a manual frequency search. The sound source was the STL ionophone (6) whose sine-wave output was varied in frequency to identify resonance peaks. It was mounted on a tight fitting piston whose position in the tube could be continuously varied. Special care was taken to achieve an airtight seal between the ionophone rod and the tube. A microphone attached to an oscilloscope was mounted close to the open end of the tube.

As a plasma is generated between the ionophone electrodes, heat is dissipated. The temperature of the air inside the tube was continuously monitored by two thermocouples, one placed near the tube opening; the other was inserted through the piston at 2 cm above the source, off-center. To keep temperatures steady, air circulation was achieved by having a fan inject a laminar air stream into the tube.

The data set for the measurements were as follows: Overall tube lengths: 12 – 21 cm; Notch depths: 1, 2, 3, 4, 5 cm; For every notch depth, measurements were also made of two tubes without the notch: The full tube length (column 1 in Fig 1) and the short tube (column 4 in Fig 1).

Figure 2 presents the plot format used to analyze the data. The temperatures recorded during the experiments were used to adjust all the frequency measurements so as to reflect a temperature of 35° C inside the tube.

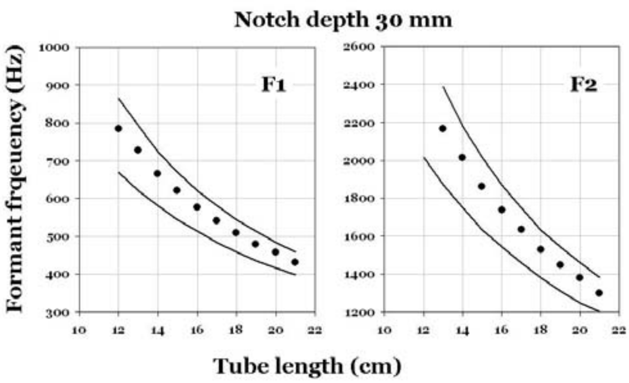

Figure 2. F1 and F2 measurements plotted against total tube length. Notch = 3 cm. The bottom and top lines represent the formant values for the short and the full-length tubes without a notch respectively.

All formants presented patterns qualitatively similar to those of Figure 3 regardless of the geometric parameter values (11).

3.2 Modeling the data

Acoustic theory enables us to predict what the values for the tubes without notches ought to be. Since the unnotched tubes have uniform cross-sectional areas, their resonance frequency would be expected to be given by:

$$f_n=(2n-1)*c/(4*(l+l_{end})) \quad (1)$$

where n is formant number, c speed of sound and l_{end} is the effective length of the pipe. At 35° C the speed of sound in cm/sec is:

$$c= 33145*(1+35/273.15)^{1/2} \quad (2)$$

We should note that the effective length of a pipe that is unflanged is given by

$$l_{end} = l + 0.6*r \quad (3)$$

where r is the cylinder radius (7).

Is it possible to equate the acoustic properties of a given notched tube with those of an unnotched tube whose length is chosen so as to match the unnotched formant pattern?

For each measured formant value we calculated the following number:

$$l_{inc} = (35204/(4*f_{obs})) - (l_{short} + 0.6*r) \quad (4)$$

where 35204 is the speed of sound at 35° C, f_{obs} is the measured formant value in Hz, l_{short} is defined as the distance (in cm) between the bottom end and the "mouth corner" of the notch. The assumption underlying the formula is that the notch will add an increment (l_{inc}) to the length of the unnotched tube (l_{short} + $0.6*r$).

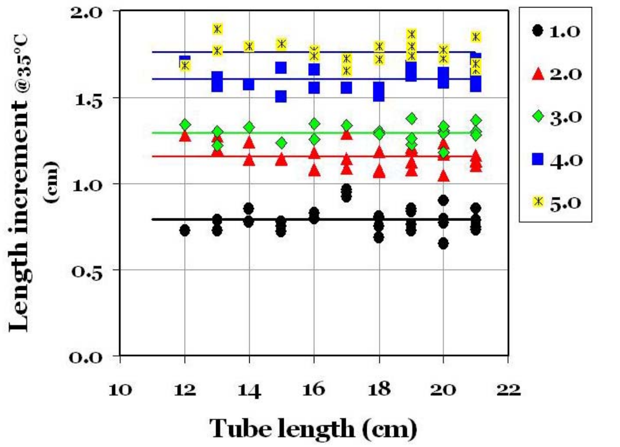

*Figure 3. Effect of notch depth on the length increment (Eq 4). When this increment is added to an unnotched tube whose effective length is [l_{short} + 0.6*r] a perfect match is obtained with the formants of the notched condition.*

Figure 3 presents an example of how the length increment varies for a tube whose total length is plotted on the abscissa and whose radius is 1.25 cm. Each colored cluster refers to a given notch depth (1–5 cm). Since formant number did not affect the increment much, formants for individual conditions are pooled. The mean is indicated by a horizontal line.

Similar results were obtained for all cross-sectional areas. For all conditions the length increment was found to increase in a lawful manner with notch depth and to some extent with tube radius.

In conclusion it would seem that the formant patterns of notched tubes can indeed be described in terms of unnotched tubes of equivalent lengths. Accordingly it is possible to decide where *"Where do notched tubes end acoustically?"*

3.3 HACMOD–a 3-D physical replica of the front cavity

How do we apply what we just learned about copper tubes with uniform cross-sectional area to vocal tracts with non-uniform cross-sectional areas? Lacking a straightforward answer we decided to try a different approach.

It would be ideal to have a physical model of the vocal tract capable of representing pharyngeal and oral cavities realistically and allowing, at the same time, the experimenter to observe the acoustic consequences of systematic manipulations of the geometry of the front cavity. In partial attainment of this goal we developed HACMOD. This tool is a hybrid device combining the use of plexiglass washers (8) with a 3-D replica of the front part of the vocal tract (Figure 4).

The replica was constructed by first making impressions and casts of a human subject's anterior oral anatomy. From these casts, acrylic models of the jaws were produced. Figure 4 presents the components of HACMOD. On the left: the maxilla with the upper teeth and hard palate. It is fastened to the metal plate and whose vertical position can be adjusted along the slits to simulate variations in jaw opening.

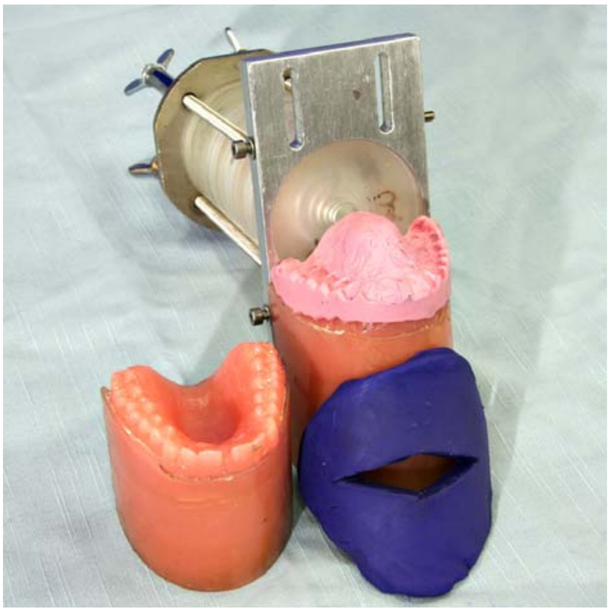

Figure 4. The HACMOD replica of the vocal tract with upper and lower teeth/jaws, lips, tongue shape and the posterior plexiglass unit.

A tongue made of lab putty is fitted into the mandibular component. A lip configuration made of play dough is also shown. In the background: HACMOD's posterior region with the pile of 0.5 cm circular plexiglass washers with variable axial holes.

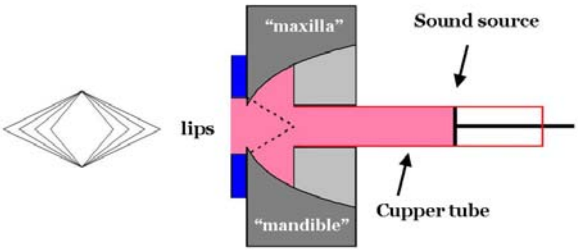

Figure 5. A resonator system [pink] formed with HACMOD as the front end [dark gray] and a copper tube as the back cavity. Opening area of lips [in blue] was varied by keeping the midsagittal separation between the lips constant while changing the anterior-posterior location of the mouth corner (as indicated by dashed lines). In this way different degrees of 'lip spreading' could be simulated.

We will report two HACMOD experiments on configurations resembling vowels with 'spread lips'. In the first, HACMOD is used as the front cavity and a copper tube as the back cavity (Figure 5). Figure 6 shows a similar set up but the back cavity is a set of plexiglass washers.

The plan for both experiments was to determine the resonance characteristics of the physical models and to compare those results with formant frequencies obtained from FORMFLEK.c, a program written by Liljencrants (9).

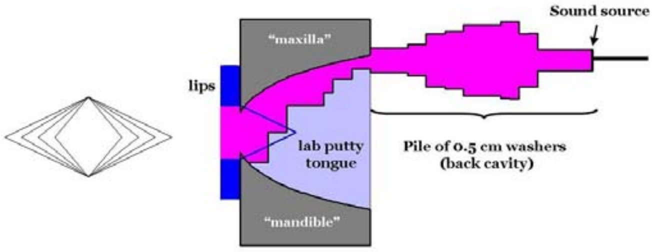

Figure 6. HACMOD and a back cavity consisting of a pile of washers set up for a 'spread' vowel. Degree of spreading was controlled by varying lip opening area as in Figure 5.

This program inspects each area function for abrupt transitions from a narrow to a larger area and applies an internal length correction l_{ilc} wherever the more constricted area is smaller than 0.16 times the larger area.

$$l_{ilc} = 0.48(A)^{1/2} [1-1.25(A_0/A)^{1/2}] \qquad (5)$$

To derive cross-sectional area functions we first made impressions and casts of the front cavity space. We used lab putty, a material developed for odontological work. Its density was established experimentally by measuring the mass and volume of chunks of this material. The front cavity casts were cut into cross-sectional slices of uniform thickness perpendicular to the vocal tract midline. These pieces were then weighed to determine their volume from their weight and density and also to derive their average areas from their volume and thickness for use in the area functions. This method implies that all area functions represented the actual volumes of the front cavities implicitly but highly accurately.

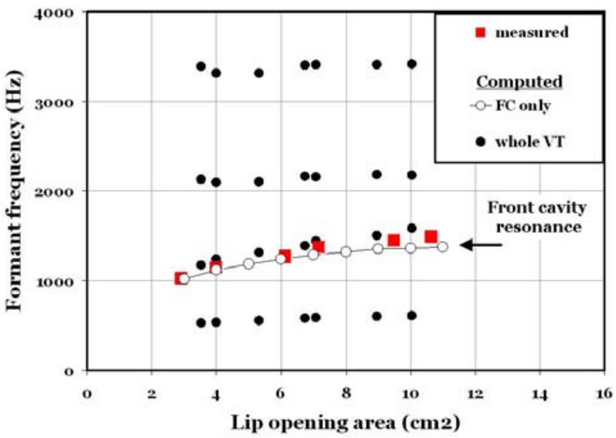

Figure 7. Measured and calculated resonances for the configuration in Figure 5. The diagram shows the effect of lip spreading on the formants of the configuration [red squares] and on values computed on the basis of front cavity volume estimates.

Figure 7 shows how formants pattern as the size of the lip opening is varied. Measurements are presented for the front cavity of the physical model [red squares] and compared with computed formant frequencies for the sound source applied to the front cavity alone [unfilled circles] and for the excitation of the complete physical model [filled symbols]. The filled circles for F1 F3 and F4 group themselves along horizontal lines. F1 appears to depend on the entire 'vocal tract' whereas F3 and F4 are associated with the back cavity. Showing a dependence on lip opening we can assume that it depends mainly on the geometry of the front cavity. The F2 values form a pattern which falls reasonably close to the frequencies calculated from the area functions. In a first approximation the agreement is satisfactory.

Let us turn to the other experiment which uses physiologically slightly more realistic cavity shapes (Figure 6). For the back cavity the area function was determined by a set of washers forming a tight seal with the HACMOD front end. A lab putty tongue was placed in the mandible producing a narrowing near the back end of the replica. Degrees of ‘lip spreading’ were simulated as before by varying the anterior-posterior location of the mouth corner while keeping the midsagittal separation between the lips constant. The front cavity volume was estimated from a lab putty cast. The results are presented in Figure 8.

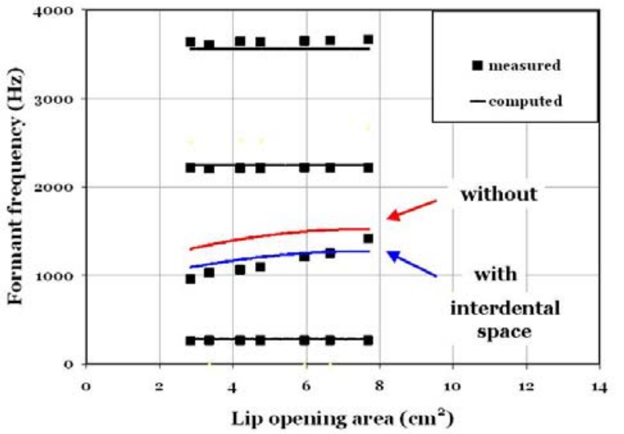

Figure 8. Measured and calculated resonances for the configuration in Figure 6. The diagram shows the effect of lip spreading on measured formants and on values computed on the basis of front cavity volume estimates. Note the role of the space between the upper and lower teeth.

Like in Fig 7, Figure 8 indicates that F1 F3 and F4 arrange themselves along horizontal lines with F1 dependent on the entire system and F3 and F4 primarily on the back cavity. As in the previous experiment, F2 increases as the lips become more spread. The calculated lines are reasonably close to observed values. In particular note the blue and the red curves for F2. They show the effect of including or excluding the space between the upper and lower teeth. The data suggest that the interdental space is relevant and that in this exercise its volume was slightly underestimated. Also we can note that the contour of the data points does not quite conform with the shape of the predicted relationship. As lip spreading increases, data points differ more from the blue line possibly indicating that the interdental space contributes less and less as the lip opening becomes larger.

The two experiments demonstrate a relatively good match between observed and predicted values which would seem to indicate that the volume-based strategy is on the right track. Note that instead of trying to find an answer to ‘Where does the vocal tract end?’ we took an indirect route. We measured the total volume of the front cavity and divided it into cross-sectional sections whose equivalent areas we derived by dividing the volume of each section by its thickness. We then plugged those areas into the area function along with the corresponding thickness of each section to obtain their length coordinates. The final section of the area function, the lip section, was specified by its thickness (length) and opening area. The answer given by the proposed procedure is then that the vocal tract terminates at the end of the lip section.

4 Effect of raised tongue blade: Exploring a twin-tube physical model

Consonants produced with an elevated tongue tip offer another challenge for acoustic phoneticians since raising the tongue tip creates a cavity under the blade. How are such articulations to be modeled acoustically? What is the most relevant way of representing this subapical space in the area function? This is a topic that has attracted some attention in the literature (12). Like spread lips this issue can be illuminated by the creation of a physical model of the vocal tract.

We chose to investigate this problem by constructing a twin-tube model (Figure 9) and applying our sweep-tone method to it.

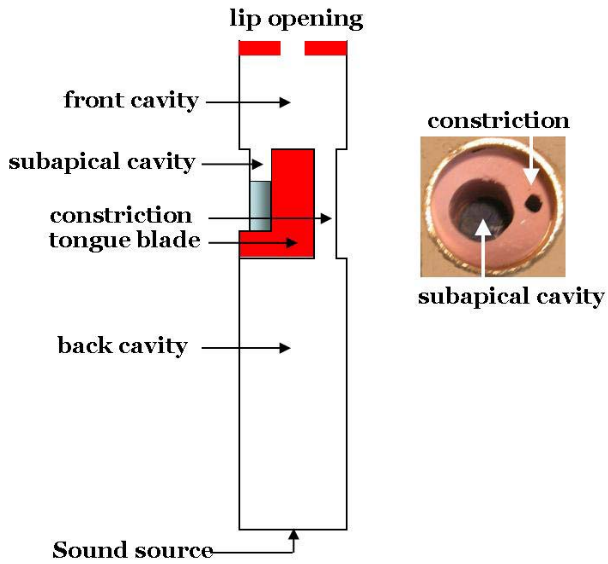

Figure 9. Twin tube model of articulations with raised tongue blade. Note that the constriction and the subapical cavity had diameters as shown on the right not as suggested in the stylized drawing.

The lip section is realized by a single washer whose cylindrical opening area is chosen from a range of values.

The front cavity is a section of a 32 mm copper tube. The length of this segment can be varied. It is tightly sealed (with duct tape) to the back cavity which is also a piece of the 32 mm tube but longer (14 cm). Tests for leakage were routinely done by inhaling through the top of the tube while holding it sealed tight against the face.

At the top it has a lab putty plug, 3 cm long, with a cylindrical constriction, 3 cm in length and with a 0.5 cm diameter. This plug also contains a ‘subapical cavity’ 2.4 cm deep but whose length can be varied by (partially or completely) filling it with water and measuring its new depth. Figure 9 presents a view looking down into this component.

The subapical space is closed at its lower end so that it forms a ‘tip and blade structure’ [in red] together with the constriction. This is a crude, but we hope instructive, way of studying the acoustic properties of coronal articulations

We shall report on two sets of measurements. In the first experiment we chose to vary the lip area and the depth of the subapical cavity. In the second study we used a small lip opening area and varied both the depth of the subapical cavity and the length of the front cavity segment.

A sample of the results from the first experiment is presented in Figure 10. The diagram uses the format of the Figure 7 and Figure 8: Formant frequencies plotted against lip opening area. The resonance frequencies for the set up in Figure 9 are shown for a completely empty subapical cavity and varying lip aperture.

The formant values for F1 F2 F4 and F5 of the twin tube are indicated by solid black circles. Computed numbers are plotted as light blue circles. As can be seen the predicted frequencies match the observed with very high accuracy the small blue symbols falling on top of the black dots.

The measured F3 results are displayed differently to highlight the behavior of the front cavity. They are indicated by solid red squares. The corresponding computed values are indicated by light blue circles.

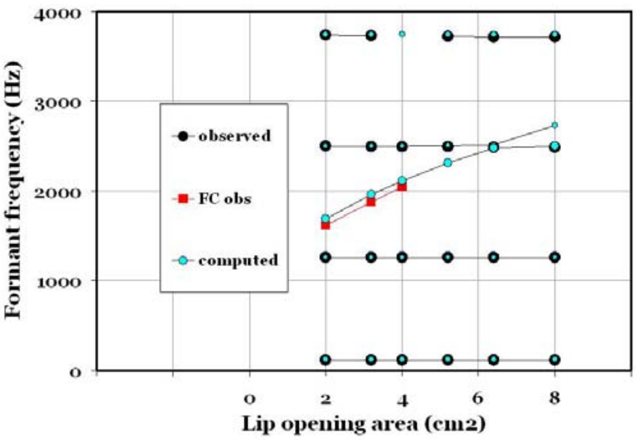

Figure 10. Formant frequencies plotted against lip opening area for an empty subapical cavity and varying lip aperture. Measured F1 F2 F4 and F5: solid black circles. Computed (all formants): light blue circles. The measured F3 (front cavity): solid red squares.

Also for F3 we note that there is good agreement between observed and computed values.

How were the predictions derived in this case? Again we adopted the volume-based approach.

Distance from glottis (cm)	Area (cm^2)
14 (back cavity)	8.04
17 (constriction)	0.20
18.1 (front cavity)	12.21
18.6 (lip section)	2.00 – 8.00

Table 1: Tube lengths and cross-sectional areas for generating the data plotted as "computed" in Figure 10. "Distance from glottis" refers to the end of each section. For the lip section the range of area values is given.

Let us give the details of the area function derivation (Table 1). The left column of Table 1 specifies the section's endpoint. Accordingly the constriction was 3 cm long; the front cavity 1.1 cm in length. In the right column the areas for the back cavity and for the constriction are simply those of their cylindrical tubes.

In this experiment the empty subapical cavity had a volume of 4.59 cm^3. Combining this volume with that of the front cavity section which was 1.1 cm long contributes an area of 4.59/1.1 cm^2. In other words the area entry for the front cavity becomes: 8.04 + 4.17 = 12.21 cm^2.

The calculations indicate that the front cavity resonance is expected to cross the back cavity formant which stays steady at 2500 Hz. In this experiment we were often unable to locate a front cavity formant for larger values of lip spreading. Or a resonance would be suggested in reponse to the sweep tone but it turned out to have an extremely low Q value and was immeasurable.

In the second experiment the lip opening was kept fixed (area = 1 cm^2) while the size of the front cavity was varied. To further explore the role of volume in tuning the front cavity resonance we present pooled data from experiments in which three front cavity lengths were used: 1.1, 0.5 and 0.3 cm. For each front cavity condition, the depth of the subapical cavity was varied by filling it with water by varying degrees. On the abscissa "Total front cavity volume" stands for the sum of the front cylinder and the subapical volumes. In deriving the area functions the method described in connection with Fig 10 was adopted. In other words, the area for the front cavity section was augmented with a contribution derived from the subapical volume.

Figure 11 shows that there is good agreement between observations and predicted values.

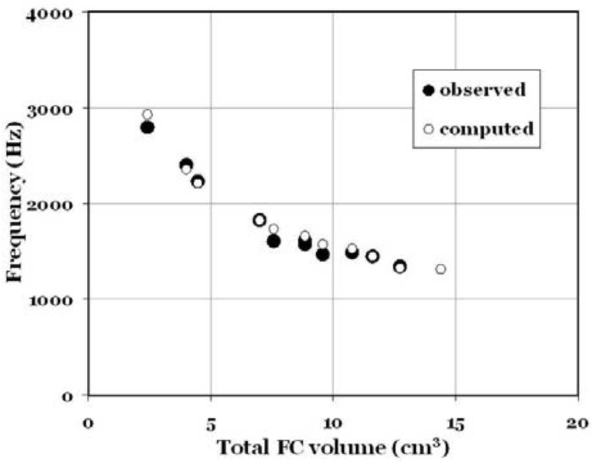

Figure 11. Formant data for the configuration shown in Fig 9 and from calculations. A constant lip opening area was used. The front cavity length was adjusted to produce volumes of approximately 2, 4 and 9 cm^3. The volume of the sublingual space ranged between 0 and 5 cm^3. Along the x-axis the sum of the front cylinder and the subapical volumes is indicated.

5 Discussion

Our report started out by identifying two unresolved problems in the acoustic modeling of the vocal tract: *spread lips* and *raised tongue blades*.

Previous research on the modeling of the lips has raised the question "Where does the vocal tract end?" (10). In this paper we approached the topic of spread lips with the aid of notched and unnotched cylindrical tubes. We demonstrated that it is possible to equate the acoustic properties of a given notched tube with those of an unnotched tube whose length is chosen so as to match the unnotched formant pattern. In other words, for cylindrical notched tubes an acoustic termination point can be specified as a length correction for an unnotched tube.

We then moved on to more speech-like geometries examining the acoustic properties of a 3-D replica of the front end of the vocal tract. Here we were forced to choose a different approach. Investigating several different front cavity configurations we were able to conclude that good formant frequency predictions were possible by taking the volume of the front cavity into account. We found that for this to work well the volume of the front cavity should include the spaces between the upper and lower teeth and the cavity that develops under the tongue blade as the tongue tip is raised.

Is there some way that we can reconcile the termination point and volume perspectives? Let us for a moment return to the copper tubes and the results on the equivalent length increments.

If cavity volume is acoustically relevant in the case of non-uniformly shaped cavities should it not be applicable also in the case of cylindrical notched and unnotched tubes?

We address this question in Figure 12. Along the ordinate we plot the length corrections observed for a given notch and tube diameter. The numbers were averaged over formant number since formant–dependent variability was found to be small and unsystematic.

The abscissa has front cavity volume that is here computed as

$$V_{FC} = hd^2/3 \quad (6)$$

where h is the notch depth, d the tube diameter. This volume is illustrated by the black areas in the insert below.

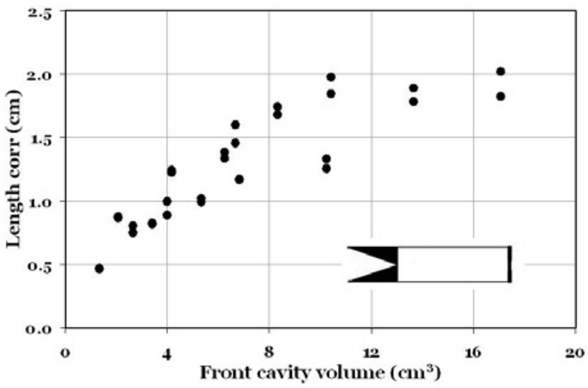

Figure 12. Data pertaining to the study of notched copper tubes. The length corrections for unnotched cylindrical tubes [that generate the formant patterns of the notched tube] are plotted on the y-axis. The volume enclosed by the tube in the notch region is on the x-axis and is illustrated by the blackened 'duck bills' of the tube.

Figure 12 demonstrates that the volumes and length corrections are clearly related, a finding which further underscores the important acoustic role played by the front cavity.

6 Conclusion

Richard Feynman, physicist and Nobel laureate, is reported to have had the following motto scribbled on his office blackboard: "*What I cannot create I do not understand"* (13).

By developing the source-filter theory Gunnar Fant was able to recreate speech and thereby met the criterion of scientific achievement that Feynman set for himself.

In a wider philosophical context the source-filter theory goes beyond acoustics and speech synthesis and analysis. It sends a strong message about scientific method. It tells us that understanding something scientifically is being able to recreate what one observes.

Few can hope to emulate Gunnar in his scientific success but we are nonetheless privileged to have his legacy and the opportunity to learn from his example.

References

[1] Fant G (1960): *Acoustic theory of speech production,* The Hague:Mouton.

[2] Fant G (2004): *Speech acoustics and phonetics,* Dordrecht:Kluwer.

[3] Fant G. (2005): "Historical notes", *TMH-QPSR 47(1)*:9-19.

[4] Lindblom B & Sundberg J (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement", *J Acoust Soc Am 50*:1166-1179.

[5] Lindblom B & Sundberg J (2008): " Issues in the acoustic modeling of the vocal tract - a progress report on APEX", *J Acoust Soc Am 123(5)*:3730.

[6] Fransson F & Jansson E (1971): The STL-Ionophone: Transducer properties and construction. *J Acoust Soc Am 58*:910-915.

[7] Kinsler L E, Frey A R, Coppens A B & Sanders J V (2000): *Fundamentals of acoustics*, Wiley:New York.

[8] Sundberg J, Lindblom B & Liljencrants J (1992): Formant frequency estimates for abruptly changing area functions: A comparison between calculations and measurements", *J Acoust Soc Am 91(6)*:3478-3482.

[9] Liljencrants J & Fant G (1975) "Computer program for VT-resonance frequency calculations", *STL-QPSR 16(4)*:15-20.

[10] Ladefoged P, Wright J & Linker W (1978): "Where does the vocal tract end?", *J Acoust Soc Am 63*, Issue S1:S32-S32.

[11] Lindblom B, Sundberg J, Branderud P & Djamshidpey H (2007): "On the acoustics of spread lips", *Proceedings of Fonetik*, *TMH-QPSR,* 50(1), 13-16.

[12] Stevens K N (1998): *Acoustic Phonetics*, MIT Press:Cambridge, MA.

[13] Gleick J (1993): *Genius The life and science of Richard Feynman*, Vintage Books:New York.

[14] Fant G (1986): "Glottal flow: Models and interaction", *J Phonetics 14*:393-399.