Subglottal formants

Fant, G. and Ishizaka, K. and Lindqvist-Gauffin, J. and Sundberg, J.

journal: STL-QPSR
volume: 13
number: 1
year: 1972
pages: 001-012

http://www.speech.kth.se/qpsr
I. SPEECH ANALYSIS and SPEECH PRODUCTION

A. SUBGLOTTAL FORMANTS
G. Fant, K. Ishizaka*, J. Lindqvist, and J. Sundberg

Abstract
Measurements on the subglottal impedance carried out on Japanese trachotomized subjects have provided a set of resonance frequency and bandwidth data from which an analog network has been constructed and added to the electrical line analog LEA. Various degrees of coupling between the supra- and subglottal systems have been studied. Results are in general agreement with observations on the spectra of the aspiration phase of stop bursts. The subglottal coupling is noticeable at glottis areas greater than 0.06 cm² providing the glottis resistance is small and the source is located at the glottis. As a result of the coupling formants of the F-pattern are shifted up in frequency by small or moderate amounts and additional subglottal formants appear. These effects can invalidate the measurement of a terminal F₁, F₂, and F₃ at the instant of release of a stop burst. Similar effects are also seen in the spectra of bursts following the release of voiced stops and it is suggested that subglottal coupling is significant here also.

In normal voicing the subglottal influence is small as long as the subglottal pressure is large enough to maintain a large glottal resistance. Weak and breathy voices would show more traces of subglottal formants than normal voices.

Introduction
This report has developed from two independent sources of activity, one in Japan concerned with impedance measurements on the subglottal system, and one in Sweden on the observation of the spectral fine structure of aspirated sound segments. These studies led to the design of an electrical line analog of the subglottal system as a refinement of vocal-tract modeling. Although the model is rather crude it has contributed to verifying the origin of extra spectral peaks in aspirated sound segments.

Pole-zero analysis of side branch and sub-source coupling
One of the basic theorems of the acoustic theory of speech production is the continuity of the F-pattern F₁, F₂, F₃, F₄ etc. within speech segments and across segmental boundaries, as determined by the continuity of the vocal-tract area function. This theory is ideally suited for a vocal tract without side branches and no coupling to the subglottal system. An extension

* Guest researcher from The University of Electro-Communications, Tokyo, Japan from November 30, 1971 - February 18, 1972.
to incorporate additional cavity systems may be treated by introducing additional sets of poles and zeros in the transfer function, Fant (1960). As long as the coupling to a side branch is small one can conceive of the added set of poles as being affiliated with the side chambers alone, and their spectral effect is limited by a set of zeros which overlap the poles at complete closure. At a large degree of coupling some of the poles will have an appreciable degree of dependence on both the main vocal-tract system and the subordinate system. The ambiguity encountered when attempting to label the poles into the two categories, (1) an F-pattern associated with the supraglottal, non-nasal system, and (2) an extra set, can for practical purposes be resolved by considering a continuity of articulation towards the uncoupled state. If this movement brings a pole into the F-pattern, it is labeled accordingly. All other poles are cancelled by zeros. For synthesis purposes this convention has the advantage of continuity and of specialization of pole-zero pairs.

Spectrographic evidence

The appearance of more peaks in a spectrum sample than can be identified with the F-pattern is accordingly indicative of a coupling to subordinate cavity systems. Of course, other explanations are also possible such as peakiness in the source spectrum, overloading during the recording and processing, artifacts in sampling procedure, or simply the periodic alternation of zeros in the spectrum.

At an early stage one of the present authors was aware of the occasional appearance of extra formants in the burst segment of aspirated stops, e.g. [t] in Fant (1960), p. 189 and [atăt], Fant (1959), p. 93.

The correct tracing of $F_2$ in aspiration preceding [a] in [pə] was felt to be a problem when deriving the F-pattern characteristics of stops, Fant (1969). Fig. I-A-1 lower left, syllable [pə:] produced by subject SÖ (=Sven Öhman) shows two formants in the $F_2$-region of the aspiration, one at 1100 Hz continuous with $F_2$ and one just above 1700 Hz. A weak trace of $F_1$ is seen at about 700 Hz and the $F_3$ at 2300 Hz is also present.

Spectral sections of the [pə:] burst sampled at the very first 5 msec of the release transient and a sample of 20 msec duration 40 msec later is shown in Fig. I-A-3 and I-A-4, respectively. These were calculated by FFT and Cepstrum processing in our computer. A model spectrum derived from
a pole-zero matching with a print-out of frequencies and half bandwidths is also included in both figures. The initial burst (just barely visible in the spectrogram) shows about two poles and one zero per kHz which is to be expected if the effective length of the tract behind the source is of the same order as the supraglottal system which would conform with glottal location of the source. In Fig. I-A-4 the zero density is higher. Here $F_1$ has moved up to 702 Hz from 475 Hz in Fig. I-A-3. $F_2$ is found at 1075 Hz with 1097 Hz in the earlier sample and $F_3$ has moved to 2250 Hz from 2297 Hz. The extra peak between $F_2$ and $F_3$ attributed to the subglottal system is located at 1620 Hz compared with 1605 Hz in the earlier sample.

This extra peak does not show up in all voices we have analyzed but it is sufficiently apparent sometimes that it cannot be regarded as an incidental artifact that could interfere with our conclusions. However, extra formants outside the F-pattern occur in many other phonetic contexts, see e.g. the aspiration in [tː] and [tiː] of subjects JLj and UR, Fig. I-A-5 and I-A-6, and the [h] of [hɔːd] spoken by UR and SÖ, see Fig. I-A-7. On the whole the extra formants are less apparent in [h] sounds than in the aspiration of stops which can be attributed to a relative smaller glottal opening in the [h] sound.

However, as evidenced by Lindqvist’s measurements from fiberoptics (forthcoming) one should not apply a static view of the glottal opening. From burst release to voicing onset the glottis is in a state of rapid closing. The subglottal formants accordingly, have a greater chance of showing up in the first part of the aspiration and this is what is actually encountered, see the [pʰ] of subject BL in Fig. I-A-2 and also the [h] of subject UR in Fig. I-A-7. A specific detail of interest in the [pʰ] of subject BL is the traces of vocal-cord edge vibrations superimposed on the aspirative noise, as judged from the temporal synchrony of $F1$ and $F2$ excitation in a part of the aspiration segment.

To collect some statistics about the observable frequency locations of peaks in aspiration segments, one of the authors, Ishizaka, compiled distribution graphs from 12 speakers, CV-syllables, see Fig. I-A-8.A, B, and C. The peak locations were measured in the first part of the burst. All possible combinations of [p], [t], and [k] and V = [aː], [aː], [uː], and [iː] were studied. (This is a part of the reference spectrographic library of the Dept. of Speech Communication.) Several regularities are apparent. In all contexts
Fig. I-A-1. Spectrograms of [ba:] and [pa:]. Two display levels for [pa:]. Subject S0.
Fig. I-A-2. Spectrograms of [ba:], [da:], and [pa:]. Subject BL.
Fig. I-A-3. Spectrum section of initial part of burst in [pə:]. Subject SÖ. Pole-zero matched curve included.
Fig. I-A-4. Spectrum section of mid part of burst in [pʰə:].
Subject SO. Pole-zero matched curve included.
Fig. I-A-5. Spectrograms of [ti:] and [ta:]. Subject JLj.
Fig. 1-A-5. Spectrograms of [ti:] and [ta:]. Subject UR.
Fig. 1-A-7. Spectrograms of [hɔːd]. Subjects UR and SC.
Fig. I-A-8.A. Distribution of peak frequencies in first part of aspiration in the release of [p] followed by [a:], [æ:], [u:], and [i:].
Fig. I-A-8. B. Distribution of peak frequencies in first part of aspiration in the release of [t] followed by [a:], [ɛ:], [u:], and [i:].
Fig. I-A-8. C. Distribution of peak frequencies in first part of aspiration in the release of [k] followed by [a:], [æ:], [u:], and [i:].
independent of vowel and consonant there is a dominant frequency of occurrence of a peak at or close to 500 Hz. The similarity of the [piː] and [tiː] distributions with peaks below \( F_2 \) at 1000 Hz and at 1500 Hz and somewhat higher locations in [kiː] is an evidence against ascribing extra formants to random noise artifacts.

According to the previous analysis the 1200 Hz maximum of [paː] should represent \( F_2 \) and the 1700 Hz peak an additional subglottal resonance. However, in view of the general discussion of coupling phenomena, if the glottal opening is large these two peaks could have an equal affiliation to the sub- and supraglottal systems. On the other hand, in [taː] the 1000 Hz peak is definitely well below \( F_2 \) which occurs in the 1500 Hz region. Other reference points are \( F_3 \) of [taː] at 2700 Hz and the \( F_2 \) of [teː] and [peː] at 1850 Hz.

What about the corresponding voiced stops? Assuming a negligible subglottal coupling no extra formants should be seen. The release transient is often weak and the spectral peaks of the burst less distinct than in unvoiced stops. However, in part similar deviations from an ideal supraglottal model are found. One example is the energy concentration between 1 and 2 kHz of the burst in most speakers' [bɔ] spectra. The upper part of this region at about 1700 Hz lies between peaks that can be identified as \( F_2 \) and \( F_3 \), see the S\Ö [bɔ] of Fig. I-A-1, where the spectral recording gain has been turned up high enough to show the \( F_2 \) of the voiced preburst occlusion at 1100 Hz. According to the vocal-tract simulation of the various phases of lip opening performed by Fant (1969) the 1700 Hz peak is too low to be identified as \( F_3 \). It was suggested that it might reflect a rapid nonlinear rising dispersion of spectral energy at the release. However, it is sometimes also seen in the aspiratory boundary interval between an [f] and a vowel [a] and it could well reflect the subglottal resonance already discussed for the unvoiced labial stop [p]. In some instances it may also be traced into the following vowel.

More peaks than in the F-pattern may also appear in the [d] burst of the syllable [da], see the 2000 Hz peak in Fig. I-A-2. As judged by the calculations of Fant (1960) it does not appear likely that the 2000 Hz peak is the plateau between two zeros.

The subglottal impedance. Analog model

The subglottal impedance data utilized in the present study was collected by Ishizaka in Japan. Five laryngectomized subjects served as subjects.
The acoustical impedance at the breathing inlet to the trachea was measured by determining the pressure fall across a known acoustical impedance at the inlet. A typical example of resistance and reactance as a function of frequency is shown in Fig. I-A-9 and the composite data of resonance frequency and 3 dB bandwidths are shown in Fig. I-A-10. Mean values of resonance frequencies are $f_1 = 640$ Hz, $f_2 = 1400$ Hz, $f_3 = 2150$ Hz, and $f_4 = 2850$ Hz. Bandwidths are of the order of 200-400 Hz and show much greater individual variations than resonance frequencies.

At this point it should be noted that the Ishizaka impedance data depart considerably from the data reported by van den Berg (1960), see Fig. I-A-10. From measurements on the cadaver of a large dog van den Berg found resonance frequencies at $f_1 = 302$ Hz, $f_2 = 870$ Hz, $f_3 = 1425$ Hz, $f_4 = 1700$ Hz, and $f_5 = 2500$ Hz. His analog model adjusted to fit human anatomical data gave similar results, $f_1 = 314$ Hz, $f_2 = 890$ Hz, $f_3 = 1390$ Hz, $f_4 = 1860$ Hz, and $f_5 = 2415$ Hz. The total equivalent length of his model was 27.4 cm. A simple scale factor difference does not explain the discrepancy. Possibly the softness and shunting effect of the alveoli walls have been neglected in the van den Berg model. A greater stiffness could also be expected under post-mortem conditions. Control experiments are highly needed to resolve the discrepancy.

In the present study we shall adopt the Ishizaka data as a basis for our modeling. He first compiled an area function of the compound trachea, bronchial tree, and lobules by summation of parallel branches. The total length was 25 cm, see Fig. I-A-11. This area function was simulated on the electrical vocal-tract analog of the Dept. of Speech Communication, see Fant (1960). Compared with the measurements on human subjects the first resonance, $f_1$, came out at a too low frequency. As shown in Fig. I-A-12 a compensation was accordingly made by adding a shunt inductance $L_B$ and resistance $R_B$ in parallel with the connecting terminals of the system. This addition could according to Ishizaka be motivated by the softness and mass loading of the muscular tissues connecting the rings of the wind-pipe.*

* An alternative explanation of the high $f_1$ of the measured data could be that the shorter branches of the bronchial tree, see Fig. I-A-24, have a greater influence than what is preserved in a simple averaging of branch lengths.
Fig. I-A-9. Resistance and reactance components of subglottal impedance. Subject Ogawa. (Ishizaka data.)
Fig. I-A-10. Resonance frequency and bandwidth data of subglottal impedance. Five Japanese subjects. (Ishizaka data.) Van den Berg data included.
Fig. I-A-12. Analog model of the vocal tract with glottal source and subglottal system.
Taking into account the particular impedance level of the line analog LEA by the conversion factor of
\[
\frac{\text{electrical ohms}}{\text{acoustical ohms}} = 112
\]  
the values of \(L_B = 410 \, \text{mH}\) and \(R_B = 1 \, \text{kohm}\) were selected. The subglottal system is terminated by the capacitance of the air in the lungs which assuming a maximum volume of 5 liter is \(C_A = 30 \, \text{microfarad LEA units}\). The glottis inductance
\[
L_g = \frac{0.5}{A_g}
\]  
is varied by a standard LEA section in terms of the equivalent \(A_g\) settings in steps of a factor 2 from 0.08 cm\(^2\) to 0.64 cm\(^2\). Assuming a constant subglottal overpressure of \(\Delta P = 6 \, \text{cm} \, \text{H}_2\text{O}\) to be maintained independent of \(A_g\) we calculated corresponding values of the glottis resistance \(R_g\) from the equation:
\[
R_g = \sqrt{\frac{2 \Delta P}{A_g}}
\]  
which is the flow dependent differential resistance, Fant (1960). Expressed in acoustical units normalized to the specific impedance
\[
R_g = 0.17 \times \frac{0.64}{A_g} \cdot \rho c
\]  
or
\[
R_g = 0.17 \times \frac{0.64}{A_g} \times 41.5 \times 112 \, \text{electrical ohms}
\]  
\(R_g = 2800 \, \text{ohms at} \, A_g = 0.08 \, \text{cm}^2\) and \(R_g = 700 \, \text{ohms at} \, A_g = 0.64 \, \text{cm}^2\). The standard value for high impedance termination of LEA without subglottal system was \(R_g = 100 \, \text{kohms or} \, 20 \, \rho c\).

The input impedance to the analog subglottal system adopted for our LEA simulations is shown in Fig. I-A-13. It matches the Ishizaka data fairly close. The maxima are found at 600, 1400, 2050, 2800, 3450 Hz and minima at 850, 1620, 2350, 3050, 3700 \ldots Hz. It should be noted that the model lacks increasing damping of higher formants found in the human data impedance.
Fig. I-A-13. Input impedance to our electrical analog of the subglottal system.
One means of testing the representability of this model is to simulate the vocal-tract sinewave sweep measurements of Fujimura and Lindqvist (1964). Fig. I-A-14 shows the transfer characteristics they obtained for a rounded open front vowel [æ] of $F_1 = 500$ Hz, $F_2 = 1600$ Hz, and $F_3 = 2200$ Hz articulated with open glottis and placing the driving source externally at the throat just above the level of the glottis in one experiment and 3 cm below the glottis in another experiment. The microphone was held just in front of the lips. In our LEA simulations of this experiment the high impedance constant current source was applied through a 100 kohms resistance at the supraglottal terminal and at the subglottal side, respectively.

The simulations provide a fairly good fit to the sweep data, as can be seen by comparing Fig. I-A-14 and I-A-15. The subglottal zero-pole and pole-zero pairs at 700-800 Hz and 1400-1500 Hz respectively, constituting a spectral distortion between $F_1$ and $F_2$, are found at identical frequencies in the simulated and measured data. An even better overall fit had been obtained if the source had been given a -6 dB per octave correction corresponding to an integration through the throat mass in the Fujimura-Lindqvist measurements. With a subglottal source, 3 cm below the glottis, there is a good general fit up to 1500 Hz as evidenced by the pronounced zero at 900 Hz in the simulation, compared to 1000 Hz in the Fujimura-Lindqvist curve.

Because of the constant current source condition of these external sweep experiments zeros occur when the impedance of the subglottal system approaches zero. A true glottal source on the other hand is a constant voltage source in series with the glottis impedance element and zeros occur at frequencies of maximum subglottal impedance. One can alternatively apply Thevenin's theorem and convert the voltage source to an equivalent current source of magnitude

$$I_s = \frac{E_s}{Z_g(t) + Z_s}$$

(6)

where $E_s$ is the constant overpressure in the lungs, $Z_g(t)$ is the time variable glottis impedance, and $Z_s$ the subglottal impedance. $Z_g(t) + Z_s$ has to be applied as a shunt across the supraglottal input terminals. Eq. (6) provides a convenient source definition if $Z_g(t)$ is high and always greater than $Z_s$ which is the usual assumption for handling the ideal non-coupled state. For simulations with line analogs the constant voltage serial source is more practical and directly applicable to the general problem.
Fig. I-A-14. Sweep-frequency vocal tract response curves, after Fujimura and Lindqvist (1964).

(a) Regular response curve with the closed glottis (supraglottal source).

(b) Open glottis, supraglottal source.

(c), (d), and (e) Open glottis, subglottal source.
Fig. I-A-15. Electrical analog LEA simulation of supra- and sub-glottal external excitation at the throat in Fig. I-A-14.
Simulation of aspiration

When dealing with aspirated segments of speech the glottis impedance is more stationary than in voicing. From the initial release of an aspirated stop to the following vowel the glottal area $A_g(t)$ typically changes from $0.6 \text{ cm}^2$ to a voicing position in about 50 msec, whereas in voicing $A_g(t)$ may alternate between closure and $0.08 \text{ cm}^2$ peak value in a few milliseconds. Stationary simulation experiments thus have a greater significance for aspiration than for voicing.

We are now in a position to apply the complete vocal-tract model to the production of stops. Fig. I-A-16 illustrates the response of a vocal tract modeled to provide a lip-rounded vowel [a] under three different conditions: (1) no subglottal coupling and glottal source, (2) subglottal coupling with $A_g = 0.64 \text{ cm}^2$ and normal $R_g = 0.17 \rho c$ with source at the lips, and (3) the same with the source at the glottis. Condition (2) would correspond to the step function pressure release. Besides the low frequency peak at 620 Hz and the zero-pole pair at 2 kHz the spectrum is fairly smooth. A -6 dB per octave correction should be added for the proper source slope. Condition (3) corresponds closely to what was discussed in connection with the spectrograms and sections of the aspiration phase of labial stops, Fig. I-A-1, I-A-2, and I-A-4. It appears as if $F_1$ and $F_2$ had moved up a few hundred Hz in frequency. The "extra" formant above $F_2$ at 1650 Hz and the doubled peaked $F_3$ also belong to the picture.

The same overall features of subglottal coupling are retained when the glottis is narrowed to $A_g = 0.08 \text{ cm}^2$ as shown in Fig. I-A-17 which pertains to a more open vowel [a] than Fig. I-A-16. The effect of varying the resistance in the glottis slit is further demonstrated in Fig. I-A-18. The normal value of $R_g$ for 6 cm H$_2$0 overpressure and $A_g = 0.08 \text{ cm}^2$, $R_g = 1.4 \rho c$ provides a large damping of $F_1$ and $F_2$ which merge to a single peak. On the other hand, a further increase of $R_g$ will gradually remove the damping in a direction towards the high resistance reference condition with effectively no subglottal influence. Fig. I-A-19 shows how an increase in $A_g$ to $0.16 \text{ cm}^2$ enhances the "extra formant" under conditions of low $R_g$ and shifts it to a higher frequency and that this is even more apparent at $A_g = 0.64 \text{ cm}^2$.

With what right may we label the first two peaks of the aspiration spectrum as $F_1$ and $F_2$? Following the general arguments presented earlier in this article we should make a check by observing what happens when the coupling is decreased.
Fig. I-A-16. Vowel [a] LEA simulation of source at glottis and at the lips.
Fig. I-A-17. Vowel [a] LEA simulation. $A_g = 0.08 \text{ cm}^2$. Small values of $R_g$. 

- **A**: $A_g = 0.08 \text{ cm}^2$, $R_g = 20 \text{ pc}$
- **B**: $A_g = 0.08 \text{ cm}^2$, $R_g = 0$, $R_{sgl} = 0$
- **C**: $A_g = 0.08 \text{ cm}^2$, $R_g = 0.34 \text{ pc}$
Fig. I-A-18. Vowel [a] LEA simulation with $A_g = 0.08 \text{ cm}^2$. $R_g$ values representative of peak values of glottal flow.
Fig. I-A-19. Vowel [a] LEA simulations with $A_g$ and $R_g$ variations.
Increasing $R_g$ brings the two peaks back to the ideal $F_1$ and $F_2$ locations as pointed out. An even more interesting procedure is to study the low resistance or loss-less case at a small $A_g$. A derivation of the particular pole-zero pattern of Fig. I-A-17, B with $A_g = 0.08 \text{ cm}^2$ is undertaken in Fig. I-A-20. The subglottal reactance including glottis is plotted as dashed lines of positive slope and the negative of the supraglottal reactance is plotted as a curve of falling slope. The intersections between these two sets of curves define the pole frequencies and the infinity points of the subglottal system define the zeros.

As the glottis area decreases, $L_g$ and thus the slope of the subglottal reactance curves increases. Pole locations will be shifted either to the infinity points of the supraglottal impedance in which case they belong to the F-pattern or to the infinity points of the subglottal system in which case they eventually overlap with the subglottal zeros and accordingly are identified as subglottal. This analysis reveals that the zero-pole pairs at about 600 Hz, 1500 Hz, 2100 Hz, and 2850 Hz are subglottal whereas the poles at 900 Hz, 1250 Hz, and 2500 Hz are supraglottal. The limiting values of these formants at complete decoupling are 720 Hz, 1080 Hz, and 2430 Hz respectively in accordance with Fig. I-A-17. Even with as small values of glottal opening as $A_g = 0.08 \text{ cm}^2$ valid for the construction in Fig. I-A-20 the shift of $F_1$ and $F_2$ is thus appreciable. Of the subglottal formants it is the 1500 Hz peak which is most apparent whilst the lowest subglottal pole is almost completely neutralized by its zero.

If the vocal tract configuration is changed to that of a homogeneous tube with liprounding superimposed the aspiration spectrum displays a more even distribution of peaks, as seen in Fig. I-A-21. This curve has some resemblance with the spectrum section of subject SÖ [pə] sampled at the release phase.

The effect of adding subglottal coupling to a vowel [i:] is illustrated by Fig. I-A-22. Because the first subglottal resonance now is well above $F_1$ both appear. The next subglottal pole-zero pairs at 1450 Hz and 2050 Hz are apparent. $F_2$ has moved up to 2400 Hz because of the proximity of the uncoupled $F_2 = 2100 \text{ Hz}$ to the subglottal pole.

A finite influence from the subglottal system on the [i] spectrum is observed down to a glottal opening as small as 0.08 cm$^2$, see Fig. I-A-23. The [i] prototype of this sample was slightly different from that of Fig. I-A-23. The locations of $F_2$ and $F_3$ are not influenced by this small degree
VOWEL [a], $A_g = 0.08 \text{ cm}^2$

subglottal zeros ●

—"— poles ▲

supraglottal poles x

Fig. I-A-20. Geometrical derivation of poles and zeros from curves of glottis plus subglottal reactance (dotted rising curves) and the negative of the supraglottal reactance (falling solid lines). Vowel [a] with $A_g = 0.08 \text{ cm}^2$. 
Fig. I-A-21. Response of neutral tube with liprounding, $A_o = 0.64 \text{ cm}^2$ and $A_g = 0.64 \text{ cm}^2$ to glottal source. Normal $R_g$. 
Fig. I-A-22. Vowel [i] with $A_g = 0.64 \text{ cm}^2$ and normal $R_g$. 
Fig. I-A-23. Vowel [i] with $A_g = 0.08 \text{ cm}^2$. $R_g$ = normal and zero.
of coupling but \( F_1 \) is slightly raised if \( R_g \) is maintained low. As observed with the [æ] vowel and as can be quite generally expected the bandwidth of \( F_1 \) is not affected at very small and very high values of \( R_g \). The maximum broadening is of the order of 100 Hz for intermediate resistance values of \( R_g = 0.5-1 \text{ \ O c} \).

It should be noted that the spectrum distortion observed with subglottal coupling, e.g. the increase of \( F_1 \) and traces of additional peaks in the spectrum, has certain similarities with that encountered in nasalization.

**General discussion**

The vocal-tract simulations of subglottal coupling have provided a quantitatively verification of phenomena observed in the noise spectra of aspirated stops. Particular attention has been given the aspiration segment preceding voicing [\( p^h \)] in the syllable [pə]. The extra formant above \( F_2 \) is associated with the second resonance of the subglottal system. The process of tracking \( F_2 \) from the voicing and back into the aspiration to an estimated place of origin (locus) at the instant of release can be invalidated for two reasons: (1) it may be hard to decide what is \( F_2 \) and what is a subglottal formant, (2) even if \( F_2 \) is correctly traced it may occupy a 100-250 Hz higher frequency than with glottis closed.

Both these phenomena provide a positive bias in estimation of the terminal \( F_2 \) location. Therefore it may be hazardous to attempt to estimate the degree of consonant-vowel coarticulation in [pə] syllables from spectrographic evidence alone. On the other hand, there are reasons to believe that a [b] is closer coarticulated with a following [æ] than is [p], i.e. that the vocal tract configuration at the instant of release is closer to a neutral position in [pə], see Fant (1969). Cineradiographic studies are needed to clarify to what extent the observed \( F_2 \) transition is a subglottal coupling artifact and to what extent it reflects an actual tongue movement during aspiration. The terminal \( F_2 \) value for [pə], reported by Fant (1969), at 1750 Hz can be anything from 100-600 Hz too high.

Subglottal formants are also seen in [h] spectra, especially in the beginning of the sound, see Fig. I-A-7, speaker UR, but not so commonly as in the aspiration segments of stops. This difference is ascribed to a smaller glottal area and a larger flow in [h] sounds.
There still remains to reveal the mystery of the 1-2 kHz spike fill in of the voiced [læ], as seen in Fig. I-A-1 and I-A-2. What is the origin of the 1700 Hz component well above \( F_2 \), as seen in Fig. I-A-1? According to our simulations it is not likely to occur as the response to a lip source. It could well be the same "extra" formant as in [pæ] produced from a glottal source. In favor of this explanation is that the release spike is temporally synchronized with the phase of glottal opening whilst the voice bar strations indicate intervals of closed glottis. Even a very small glottis area provides a substantial subglottal coupling if the glottis resistance is low enough and the latter condition may be explained by the relative large amplitude of vocal fold vibrations and a finite delay in the fall of the supraglottal pressure and thus of low flow through the glottis. However, the glottal flow must be large enough to create a noise source and it is uncertain how these conditions could be combined. Further studies are needed.

One can also see traces of the 1700 Hz "extra" formant in the following vowel. The extent to which subglottal resonances influence the properties of the voice source is not clear yet. Assuming a main instant of excitation at the closing of the glottis the response cannot be influenced by subglottal phenomena until at the instant of opening and at this place the magnitude of excitation is small. On a time average basis, however, there should occur a small but finite effect. It is not probable that the 800 Hz source spectrum minimum often found in male voice spectra is related to a subglottal impedance maximum. It is more likely to reflect the waveshape of glottal pulses as suggested by Flanagan (1965).

A suggestion for further studies would be to repeat the Ishizaka measurements of subglottal impedance on a number of Western subjects. Lower resonance frequencies would be expected on account of anatomical scale factors. Several indirect methods could also be used, e.g. a sine-wave sweep with the Fujimura-Lindqvist method or analysis by synthesis of aspiration spectra. It should be noted that the peaks of subglottal impedance constitute the frequencies of minimum response, i.e. zeros in the aspiration segment and that associated poles are located above the zeros at a distance determined by the degree of coupling. The experimental data of Fig. I-A-8 on spectral peak locations are not conclusive enough but would allow for one more peak in the 500-2000 Hz region than what follows from Ishizaka's subglottal impedance model. Van den Berg's model which contains one more resonance than ours...
larynx
right bronchus
glandula thyroidea
trachea
superior lobe of r.l.
middle lobe of r.l.
basal lobe of r.l.

Fig. I-A-24. Larynx, trachea, and bronchial tree.
might after all be valid for some speakers. Further modeling of anatomical data, see Fig. I-A-24, should be made with a view to study how the distribution of bronchial branch lengths affect the impedance curve. Theoretically, an uneven distribution of lengths would account for extra pole-zero pairs.

Acknowledgment

We gratefully acknowledge the contribution of Johan Liljencrants who performed the computer analysis and spectrum matching of spoken items.

References


LINDQVIST, J. (forthcoming): "Laryngeal articulation studied on Swedish subjects"