Analysis and Validation of Higher Pole Correction Function

U.K. Laine
Stockholm, Sweden

1. Introduction

The concept of higher pole correction is connected to the all-pole modelling of speech production. Primarily it is used in the cascade terminal analog realizations where only the first few formants are included. The contribution of the higher formants to the lower frequency region is modelled by a higher pole correction (HPC) function to get the formant peaks in the spectrum to the proper levels. The modelling of HPC is usually (Fant 1959) determined by two independent parameters: the area of mouth opening and the physical length of vocal tract. The mouth opening determines the size of the radiation inductance which lowers the formant frequencies. This effect of the radiation inductance can be considered as a change in the physical length, l, to give an 'effective' length $l_e = l + \Delta l$, where $\Delta l = 0.8*\mathrm{SQRT}(A_o/\pi)$ and A_o is the area of the mouth opening. The term Δl is also called end correction. Henceforth the higher pole correction concept will be discussed only in terms of the effective length.

In the derivation of the higher pole correction function for any vowel sound it is assumed that the higher poles are located as they are in the lossless neutral vocal tract and that the higher poles have infinite Q-values (bandwidths of higher poles equal zero). Also the possible effects of cross-modes to the lower frequencies have not been discussed. The above derivation by Fant has been generally used. However, there have been no validating studies to verify the assumptions mentioned and the consequent effect on the accuracy of the HPC function.

In this study the true volume velocity transfer function is measured foı acoustic tubes of different shapes. Then the corresponding transfer functions are also calculated by a computer simulation of the transmission line (TL) model. By a comparison of the physical measurements with computer simulation results, the validity of the lip radiation impedance in particular, and the TL model in general, are studied. The computational TL model is then used as a reference system to study the HPC function in s- and z-domains.

2. Acoustical Measurements

To simulate vowel production, a straight uniform hard-walled tube of plexi-

glass was mounted into a spherical baffle with a radius of 9 cm. The tube was 17 cm in length and 2.95 cm in inner diameter (area 6.84 cm^2). The 'glottal' end of the tube was totally closed with a hard wall and the open end was mounted exactly on the surface of the baffle. Electrical sparks were used as excitation at the glottal end. The measurements were carried out in an anechoic room (Laboratory of Electroacoustics, Helsinki University of Technology). A real time spectrum analyzer was directly connected to the microphone amplifier to document the volume velocity transfer properties of the artficial vocal tract.

One measurement for the neutral tract is shown in Fig. 1. The corresponding transfer function calculated from the TL model (see sec. 3 for details) is superposed on the same figure. One can note that the match is fairly good up to about 7 kHz. One typical measurement for the nonuniform tube with the result from TL simulation is given in Fig. 2. In this case a three tube model was obtained by placing a block of 8.5 cm length into the tube, starting 2 cm from the 'lips', yielding a constriction area of 1.77 cm^2. This profile produces a response similar to the vowel /e/. Also in this case the TL simulation gives a good match to the physical measurements. These examples illustrate the validity of the TL model and the radiation impedance model used as a computational reference system in this study.

Since the TL model which describes only one-dimensional wave behavior in the tract, still gives a good match to the acoustical measurements, we can infer that the cross-modes in the measured physical model have only little effect on the lower frequencies. Thus it seems reasonable to leave out the cross-modes in the HPC derivation.

It can be noted from Fig. 2 that the formant frequencies have moved away

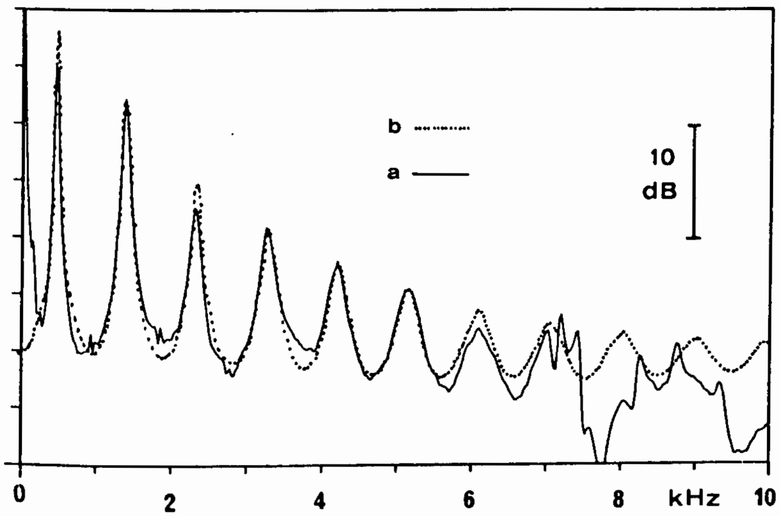

Fig. 1. Frequency response of the a) physical model b) TL model.

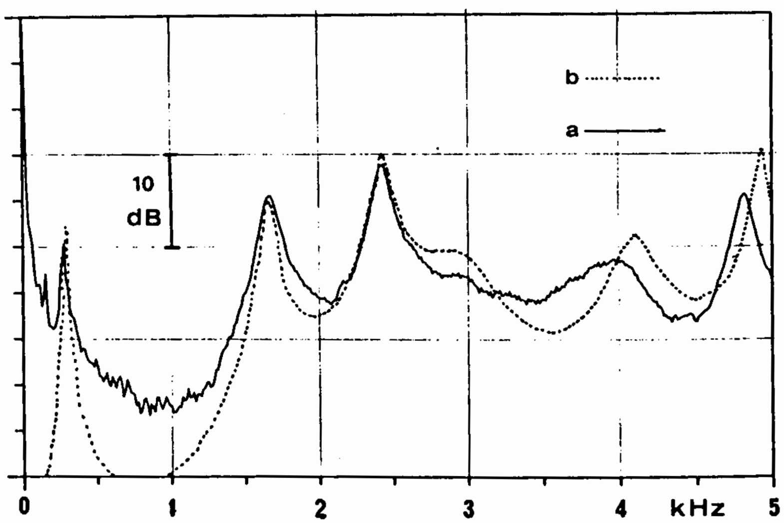

Fig. 2. Frequency response of the three tube model a) physical moaeı b) TL model.

from their neutral positions. Also, even in the case of a neutral tube (Fig. 1) the bandwidths are increasing towards the higher formants due to the larger radiation losses. How these factors affect the HPC function will be seen in the following analysis.

3. HPC Function in the S-Domain

As known, by assuming one-dimensional wave behavior in the vocal tract, its sound transmission properties can be described by applying the theories of transmission lines, The acoustic transfer functions are then given by the hyperbolic functions, which means that the model cannot directly be used for the time domain simulations. Therefore some approximations are needed to develop a corresponding lumped parameter model with rational transfer function (Skilling 1951, Flanagan 1972). However, in this study only the analysis in the frequency domain is needed and so the original TL model can be used in its full mathematical accuracy without any approximations. All frequency dependent losses (viscous friction, heat conduction, radiation etc.) can also be calculated from their original equations with the correct frequency dependency. The lip radiation impedance was approximated by a piston in a sphere model and calculated from the exact serial expansions given by Morse and Ingard (1968).

Vocal tract profile data for Russian vowels estimated from X-ray pictures by Fant (1960) were used in the TL model to calculate the corresponding volume velocity transfer functions. Values for the first five formant frequencies and bandwidths were estimated for the vowels /a/, /o/, /u/, /i/, and /e/.

Transfer functions for analog cascade resonant filters with the estimated formant parameters were calculated and substracted (in log. magnitude form) from those given by the TL model. Smoothly rising curves corresponding to the true HPC were achieved by these means. The HPC functions for the five formant model were then calculated by using Fant's formula and the known effective lengths. The error curves are shown in Fig. 3. The error is typically very small below 4.5 kHz, and then grows rapidly. This is mainly due to the sixth formant of the tract which usually has a lower frequency than assumed in the derivation of the HPC function.

The s-domain simulations confirm clearly the validity of the HPC function given by Fant. The function is seen to be an accurate and useful tool to be used in the all-pole modelling of the vowel tract transmission. In the analyzed five formant case the 3 dB point in the error curve is typically located between 4.2-4.6 kHz. Although the two assumptions made during the derivation of the HPC function are not confirmed by acoustical reality, the HPC function is seen to be quite insensitive to the absolute positions of the higher poles. Only their average density has to be known. The results are also in harmony with the fact that the bandwidth of the formant has primarily a local effect only. These general notes are also valid in the following analysis made in z-domain.

4. HPC Function in the Z-Domain

Gold and Rabiner (1968) have published a comparative study on analog and digital all-pole models. They compared the digital and analog transfer functions for ten English vowels. The models had three controllable for-

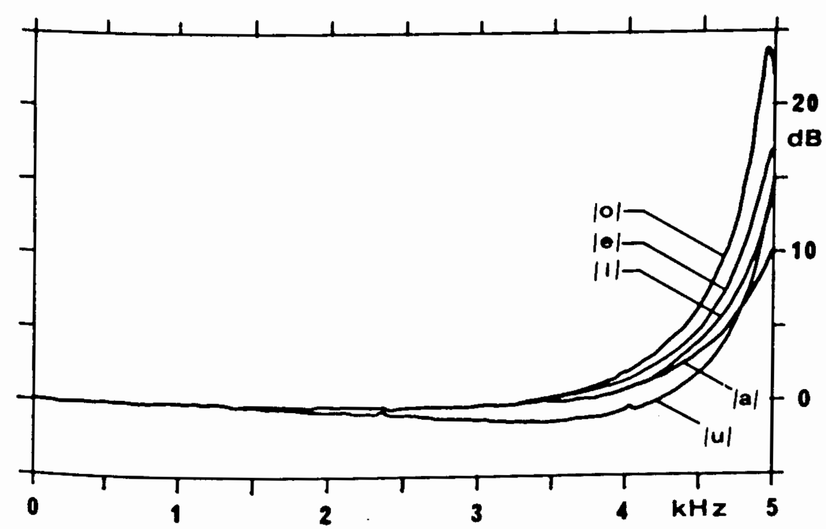

Fig. 3. Difference curves in the analog domain between the calculated, true HPC function and Fant's formula.

mants together with two or seven fixed higher poles corresponding to the 5- and 10-pole systems analyzed. As a result it was reported that the 10-pole analog system with HPC function and the 10-pole digital system without any HPC are 'extremely close' to each other, and it was also said that: 'This strongly indicates that higher poles of the vowel tract transfer function are automatically and more or less correctly taken into account by the repetitive nature of the digital formant frequency response. We also note that this intrinsic correction is actually more accurate than the quite good analog higher pole correction used in our computations. These results are generally valid for all the vowels'.

To have a closer look into this question a cascade all-pole model was compared with the TL model as in the previous s-domain study. Also five formants of the vowels were used and the sampling frequency was fixed at 10 kHz. This sampling frequency gives an optimum aliasing effect for a neutral tube of effective length 17.5 cm.

The results of the simulation are collected in Fig. 4. The difference between the TL model and the digital model is largest when the effective length of the tract differs most from the length fixed by the sampling frequency (17.5 cm). In this simulation one of the largest errors appeared for vowel /u/. This case is illustrated in more detail in Fig. 5, where the TL transfer function is shown together with the digital one. It can be noted that due to the large effective length of the actual vocal tract the digital model has approximately missed one formant (the 6th one). If we had the possibility to control the sampling frequency we could achieve a fairly good match between the models.

One can clearly see from the above results that the digital 5-pole model

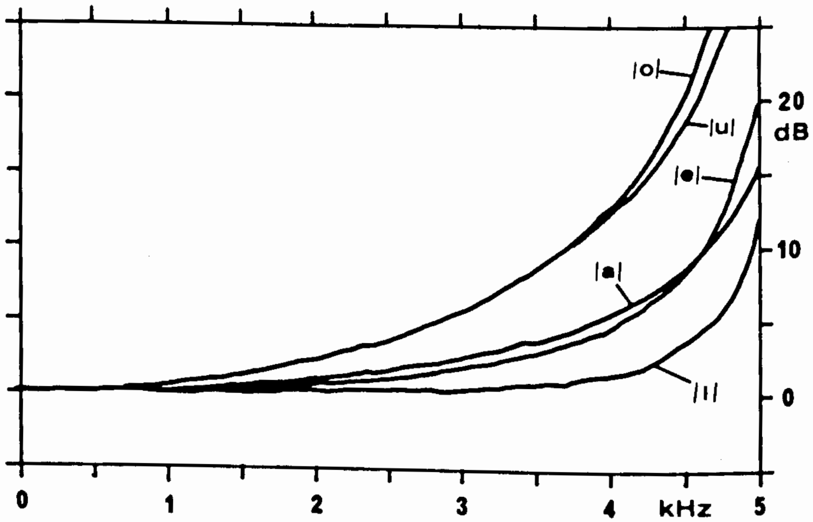

Fig. 4. Difference curves in the digital domain between the calculated TL model transfer function and the all-pole digital model.

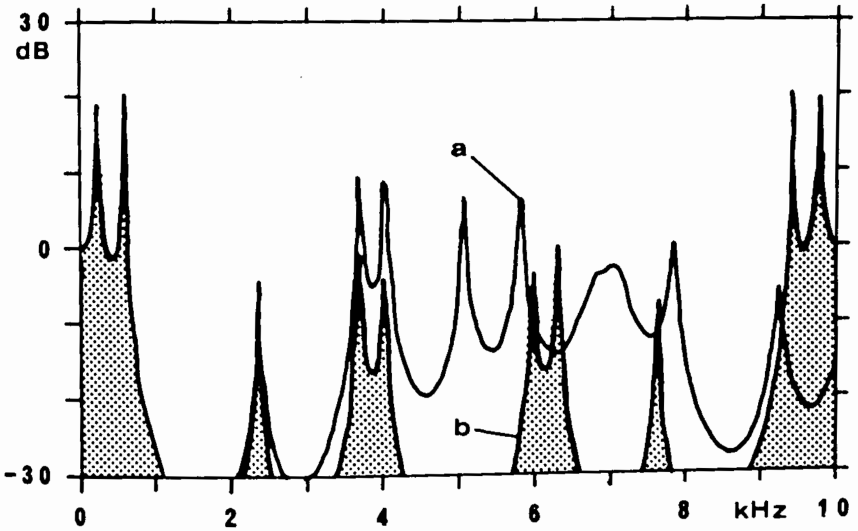

Fig. 5. Log. spectrum of /u/: a) TL model b) digital 5-pole model (aliased spectrum).

without any HPC does not give as accurate results as the corresponding analog model with HPC. Our result contradicts the conclusions made by Rabiner and Gold. However, their study was limited in the following sense: they did not compare the terminal analogs directly with the more realistic distributed parameter transmission line model because as they mentioned, it is 'quite difficult'. They have chosen some artificial values for the higher poles and compared the digital models directly to the analog one. The higher poles chosen led to vocal tract models with a fixed effective length of 17.5 cm, whereas in reality the effective length of the vocal tract varies.

Acknowledgements

The author wishes to thank Dr. M. Karjalainen at Helsinki University of Technology for organizing the acoustic measurements used in this study and Prof. G. Fant for the comments upon the manuscript. The author is indebted especially to Dr. T.V. Ananthapadmanabha for many critical, constructive comments and discussions upon this work. This study has been supported by the Academy of Finland, Nordiska Forskarstipendier and the Cultural Foundation of Finland.

References

Fant, G. (1959). Acoustic Analysis and Synthesis of Speech with Applications to Swedish, *Ericsson Technics*, No. 1.

Fant. G. (1960). *Acoustic Theory of Speech Production*, Mouton, The Hague.

Flanagan, J. (1972). *Speech Analysis Synthesis and Perception*, Springer-Verlag, Berlin, NY.

Gold, B., Rabiner, L.R. (1968). Analysis of Digital and Analog Formant Synthesizers, *IEEE Tr. on Audio and Electoac.*, **16** (1), 81-94.
Morse, P.M., Ingard, K.U. (1968). *Theoretical Acoustics*, Mc Graw-Hill, NY.
Skilling, H. (1951). *Electric Transmission Lines*, Mc Graw-Hill, NY.