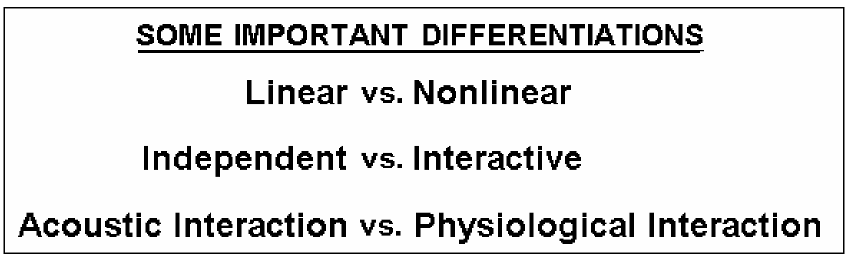
The Source-Filter Model Lives (if you are careful)

A paper presented by Martin Rothenberg at the Voice Foundation 37th Annual Symposium
May 28 to June 1, 2008

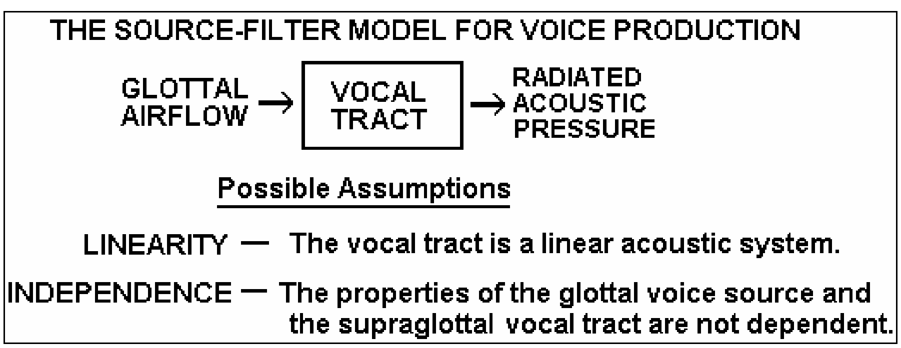
The goal in this paper was to clarify the conditions under which the source –filter model can be used in the study of voiced speech and present an explanation of the source-filter model for voice production that differentiated between the linearity and independence assumptions. In doing this, the paper outlines some of the more important conclusions from my own work in source-filter interaction, as published in papers during the period 1977 to 1988. The comments after each slide represent remarks in the verbal presentation, and in some cases may be an abbreviation of those remarks or an expansion or clarification.

SLIDE 1



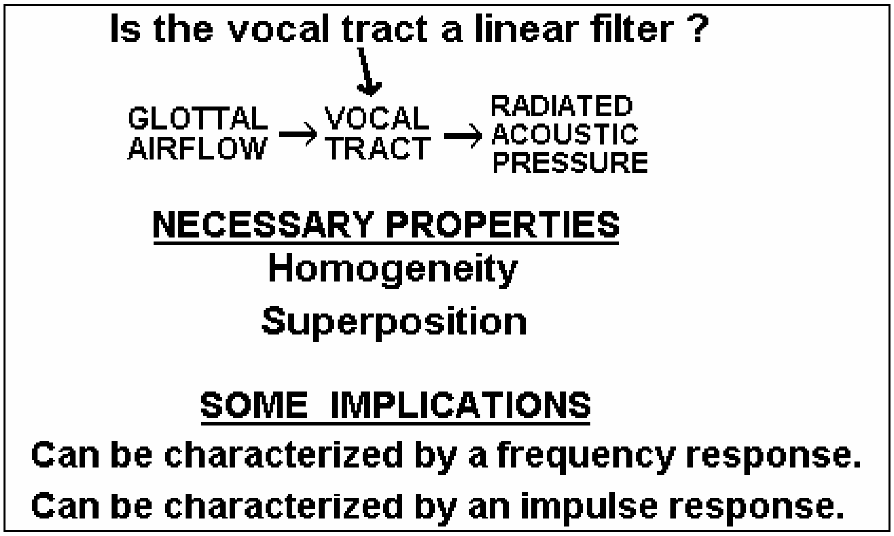
These separate and distinct differentiations must be made clear for a proper understanding of the uses and limitations of the source-filter model for voice production.

SLIDE 2



In the source-filter model, linearity and independence are separate assumptions, and either can hold without the other. Note that it is the vocal tract as an acoustic system that is assumed linear in the source-filter model, and not the glottis as a valve for airflow. In the simplest model, both properties are assumed to hold. This model is useful for pedagogical purposes, some speech synthesis applications and, more generally, as a first approximation for non-breathy voiced speech.

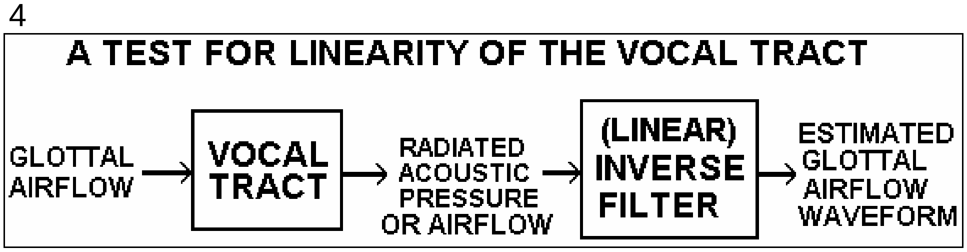
SLIDE 3



Linearity is defined mathematically for a system (or mathematical function) that has an independent variable (the input) and a dependent variable (the output). The term describes certain relations between the input and output. Without going into mathematical detail, it can be shown that a hard-walled system of tubes with no sharp bends, extreme constrictions or sharp projections into the flow path (demonstrated as a trumpet horn) is a linear acoustic system for sounds of reasonable amplitude. A kazoo (demonstrated not to follow the rule of homogeneity) is clearly not a linear system. The vocal tract is fairly linear acoustic system, if vibration of the softer walls, such as the cheeks or velum can be neglected. Physiological systems are generally <u>not</u> linear, though a linearity assumption may be useful in some applications.

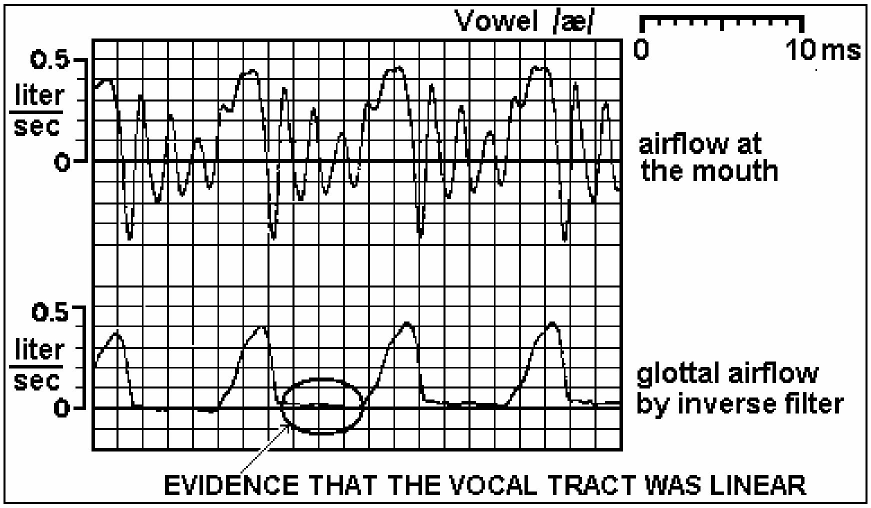
If and only if an acoustic system is linear, can it be characterized by a frequency response or the response to an impulse and the output obtained by multiplying the input Fourier spectrum by its frequency response, or by convolving the input waveform with its impulse response.

SLIDE 4



If a linear inverse filter can be derived that yields an accurate estimate of the glottal airflow waveform, then the vocal tract can be considered a linear system.

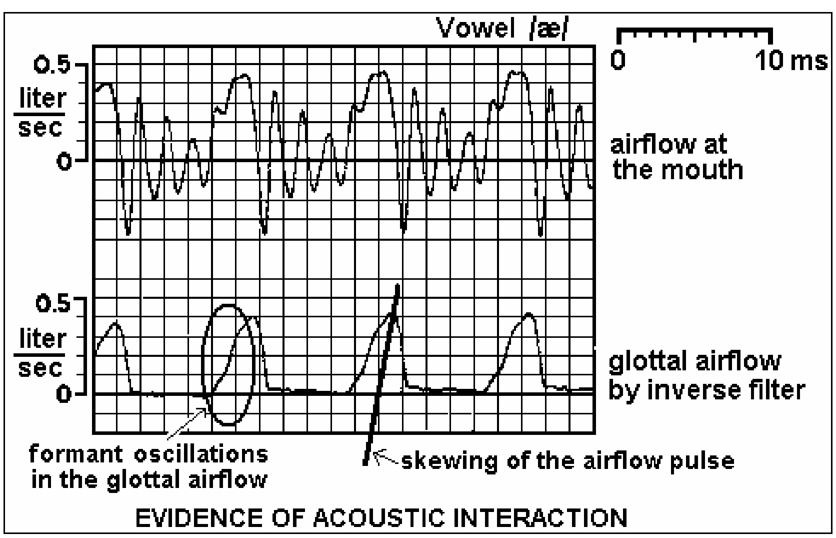
SLIDE 5



We generally do not know the glottal airflow pattern sufficiently to verify the entire glottal airflow waveform, however, for a speaker or singer known to have a clear closed glottal phase, we do know that the glottal airflow must be zero during that period. The presence of an indication of zero airflow in the inverse filter output during the closed phase can then be used as a test for linearity of the vocal tract. In the above example, slight variations that can be seen in the zero flow period indicate that there was a slight nonlinearity in the vocal tract, the airflow transducing system or the inverse filter, but the vocal tract was essentially linear.

[Slide adapted from M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing", ", *J. Acoust. Soc. Amer.* 53, 6, 1632-1645 (1973)]

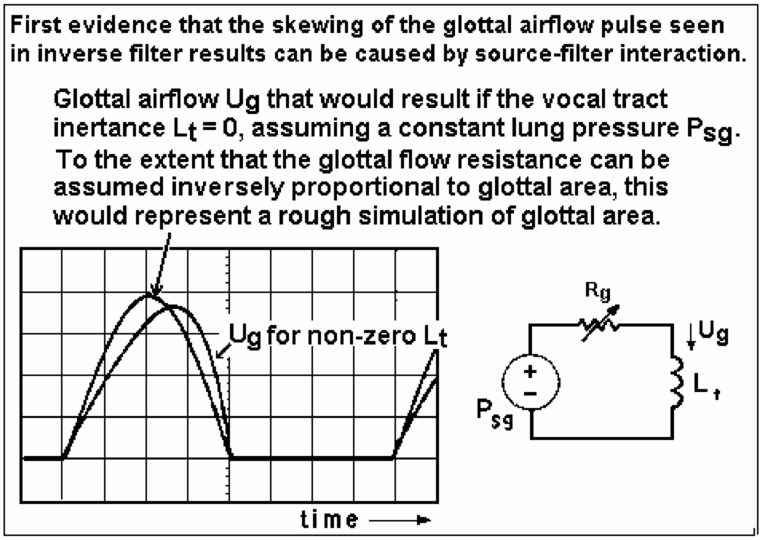
SLIDE 6



<u>Acoustic</u> interaction refers to the acoustic properties of the vocal tract affecting the airflow at the glottis (but not necessarily the vibratory pattern of the vocal folds). It was clear from the earliest inverse filter results, going back over 40 years, that at least two types of interaction between the vocal tract acoustics and the airflow at the glottis were present, as indicated in the slide.

We consider first the skewing of the glottal airflow pulse, which can cause a large increase in the energy of the higher harmonics from the energy that would be predicted if the airflow roughly followed the glottal area function.

SLIDE 7

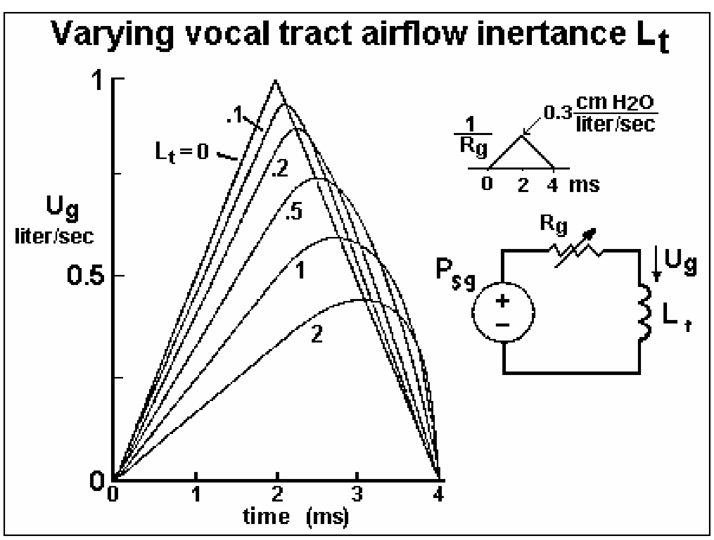


The mechanism causing the skewing of the glottal airflow pulse, especially pronounced in strong voices, was first described in conjunction with this figure taken from an electrical simulation reported in article in the Journal of the Acoustical Society in 1977. The element Lt simulated the inertance of the air in the vocal tract (and possibly the glottis and the trachea near the glottis).

An early, and possibly the first, mention of this type of mechanism occurring in a common musical instrument, the trumpet, appears to be by Backus and Hundley in the same journal in 1971. (“the mechanism primarily responsible for harmonic generation is the relationship between the input impedance of the trumpet and the time-varying impedance of the player's lip opening during a cycle”) Interestingly, it is also the mechanism that was used in the Model T Ford to generate a high voltage for the spark to ignite the gasoline-air mixture in each cylinder, with the closing of the vocal folds being analogous to the interruption of current flow in the inertive ‘spark coil’, as caused by the opening of the ‘points’ in the distributor.

[Slide from M. Rothenberg and S. Zahorian, “Nonlinear inverse filtering technique for estimating the glottal-area waveform”, *J. Acoust. Soc. Amer.* 61, 4, 1063-1071 (1977).]

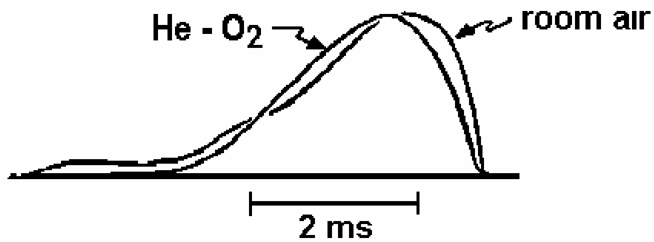
SLIDE 8



A few years later we found a solution to the differential equation for the simple model in Slide 7 for interaction with vocal tract inertance that enabled us to plot the airflow pulse waveform for varying values of vocal tract inertance, assuming a symmetrical triangular variation in glottal flow conductance (the inverse of flow resistance and roughly related to glottal projected area). The results showed a pulse waveshape similar to the waveshape found in inverse filtering very strong voices using an open vowel for an inertance value of 2, in the units used for the figure. The pulse waveshapes for inertance values of between 0.5 and 1 were similar to those that can be seen with weaker voices. The results indicated that if the inertance was high, then 1. the shape of the airflow pulse became relatively independent of the waveshape of the glottal area variation, and 2. the interaction with vocal tract inertance tended to reduce the average glottal airflow.

[Slide from M. Rothenberg "Acoustic Interaction Between the Glottal Source and the Vocal Tract" in *Vocal Fold Physiology*, K, N. Stevens and M. Hirano, Eds., U. of Tokyo Press, pp. 305-328 (1980).]

SLIDE 9

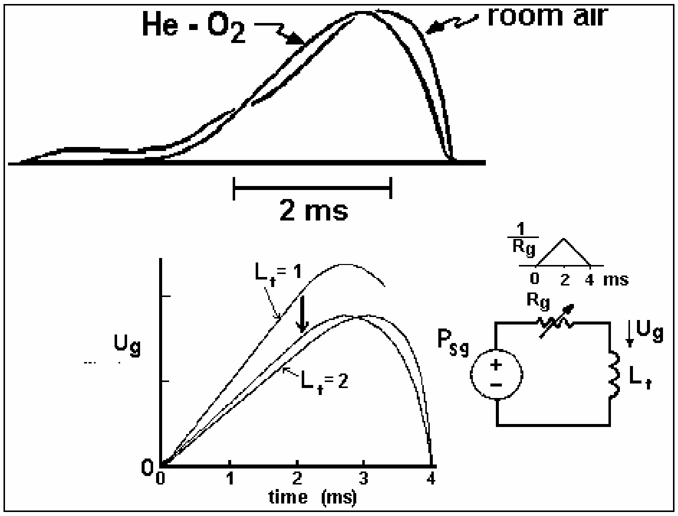


Use of Helium to reduce interaction with vocal tract inertance.

Though the evidence for skewing of the glottal airflow pulse coming from an interaction with vocal tract inertance was persuasive, we wanted to tie this down by varying the inertance and looking at the result. Changing the dimensions of the laryngeal vestibule, a likely source of much of the inertance, was deemed impractical. Instead, we tried replacing the nitrogen in the speaker's lungs with the much lighter helium, and having a trained singer vocalize in the same manner with air and then after breathing the helium-oxygen mixture. In the figure shown in the slide, the peak airflow values for the two cases have been equalized to make a visual comparison easier. (Both the average and peak airflow values appeared to be greater with the helium, however in making this measurement the wire-screen mechanism used for measuring airflow was not recalibrated for a helium-oxygen mixture.)

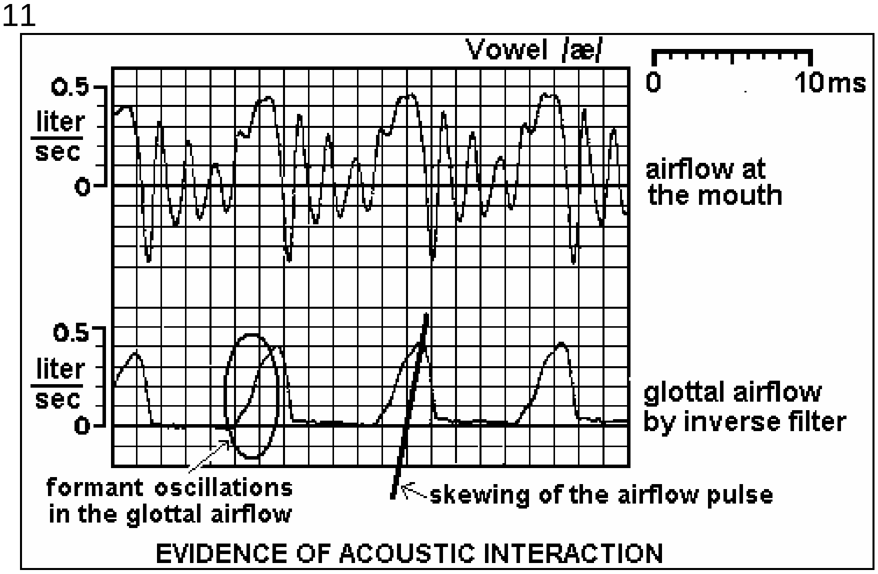
[Slide from M. Rothenberg, "Source-Tract Acoustic Interaction and Voice Quality", *Transcripts of the Twelfth Symposium: Care of the Professional Voice, June 6-10, 1983,* The Voice Foundation, New York, NY, pp. 15-31 (1984).]

SLIDE 10



In this slide, the helium-oxygen results are compared with the airflow pulse waveforms obtained from the simple model for source-tract inertive interaction. For the model results, only two inertance values are shown and the amplitudes for the two values of inertance have been normalized to expedite visual comparisons. The similarities between the model result and the result from the lowering of inertance by breathing helium are apparent.

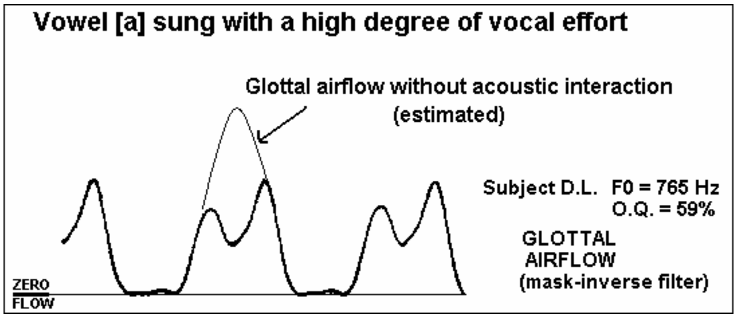
SLIDE 11



Next let us look at the acoustic interaction caused by the strong first formant energy just above the glottis. In the example in the figure, there is an oscillation at the frequency of the first formant in the glottal airflow pulse. This is not an error in inverse filtering – the oscillation represents formant energy passing from the supraglottal vocal tract into the trachea as the vocal folds separate. Note that the oscillations in the airflow at the mouth decrease more rapidly as a result.

One implication of this for the professional voice is that, other factors being equal, a larger closed quotient will result in more formant energy being radiated, at least in this pitch range. This conclusion agrees with our observations of voice airflow waveforms over the years.

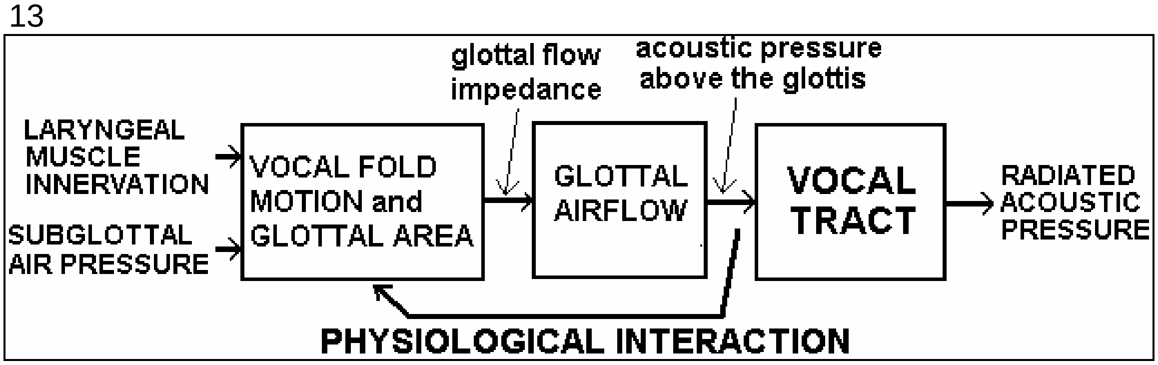
SLIDE 12



The formant interaction in the previous slide represented what is generally found in productions in which the formant frequency is much above the fundamental frequency F0. If the formant is closer to F0, the formant related acoustic interaction can be much stronger. In a 1986 paper I described the potential significance of a soprano placing the first formant near the fundamental frequency of a strongly sung /a/ vowel at F#5. (The actual F0 for the cycles shown was slightly lower than F#5 because of the vibrato used.) F1 was estimated from the inverse filter settings to be 749Hz, + or – 10 Hz, or slightly less than F0. The interaction appeared to reduce both peak and average glottal airflow, and strengthen the energy at the higher harmonics relative to the energy at the lower harmonics. These results would not be predicted from a non-interactive model.

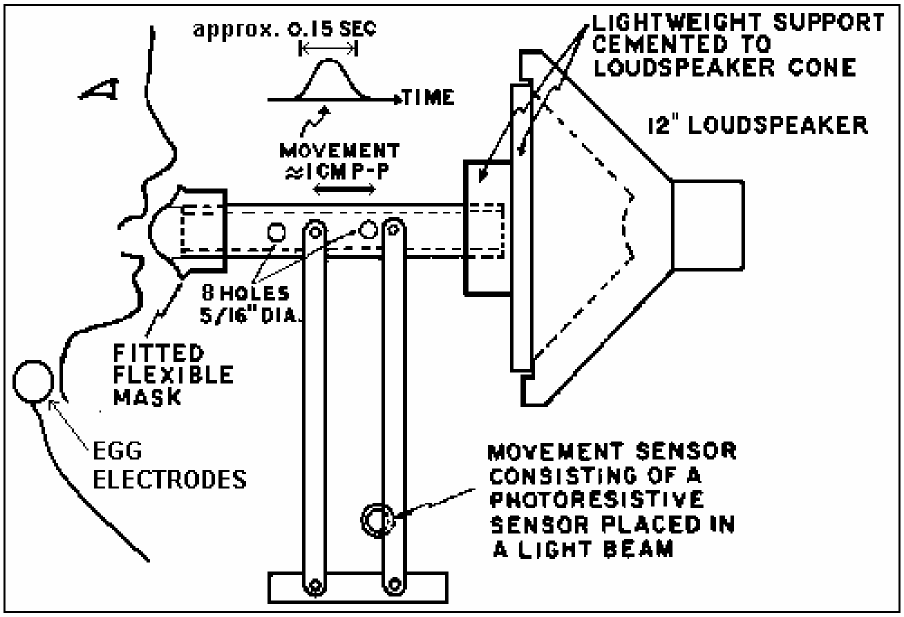
[Slide adapted from M. Rothenberg, "Cosi' Fan Tutte and What it Means or Nonlinear Source-Tract Acoustic Interaction in the Soprano Voice and Some Implications for the Definition of Vocal Efficiency**"**, in *Vocal Fold Physiology: Laryngeal Function if Phonation and Respiration*. T.Baer, C. Sasaki, and K.S. Harris, eds., College Hill Press, San Diego, pp. 254-263 (1986).

SLIDE 13



In what I prefer to call <u>physiological</u> interaction, the vocal tract acoustics affect the pattern of vocal fold vibration. Physiological interaction, as defined here, is common in musical instruments. In a harmonica, for example, the vibration frequency of many reeds can be changed by a semitone or more by an experienced player by properly shaping the mouth and pharynx chamber behind the reed. (demonstrated) In instruments such as the trumpet, the vibration frequency of the lips, which defines the note played, is a complex function of the physiological adjustment of the lips, the amount of driving pressure, and the resonances of the trumpet horn. These observations with musical instruments led us to question why there was not more such interaction reported for the human voice.

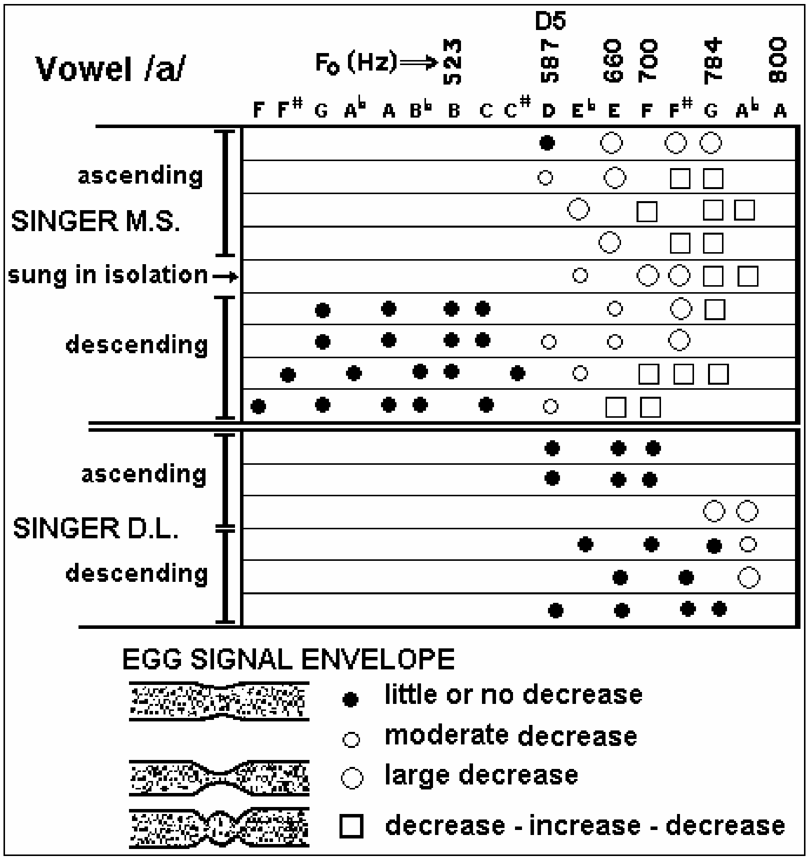
SLIDE 14



To test for a possible sensitivity of the voice to the acoustic properties of the vocal tract, we looked for conditions that may have been most conducive to physiological interaction, and chose the soprano voice at pitches high enough so that the vocal tract could easily have F1 near F0. The high pitches might also bring a marginally stable vibratory pattern that could be most easily affected by the interaction. In 1988 we reported an experiment in which the vocal tract formants were momentarily shifted during a particular note by moving a tube to be briefly in proximity to the singer's lips, thus acting to effectively extend the length of the vocal tract and lower formants. Perturbation of the vocal fold vibratory pattern during this maneuver, if it occurred, was detected by monitoring the pattern of vocal fold contact area with an electroglottograph.

[Slides 14 and 15 are from M. Rothenberg: Acoustic Reinforcement of Vocal Fold Vibratory Behavior in Singing, in *Vocal Physiology: Voice Production, Mechanisms and Functions* , O. Fujimura, Ed., Raven Press, New York, pp. 379-389 (1988).]

SLIDE 15



In brief, the results for two professional sopranos, as summarized in the slide, showed that singer D.L. had very little physiological interaction except at the highest point in the F0 range tested, near A^b6. On the other hand, singer M.S. showed significant physiological interaction at all pitches at or above D5.

These results appeared to indicate that with a more stable vibratory pattern, D.L. could be free to choose the vocal tract configuration more freely than M.S. If true, this might be an important feature of the singer's singing style. I expected that other researchers would follow up this lead and investigate the relation of voice quality to vocal fold vibratory stability in the pitch range measured in this experiment. However, that was over 20 years ago, and this has not happened to my knowledge.

It was also mentioned in the discussion that in a 1984 paper we showed that the effect of inertive interaction on <u>breathy</u> voice is opposite to the effect on non-breathy voice, and that measurements on breathy voice can be used to estimate the vocal tract inertance.

[M. Rothenberg, Source-Tract Acoustic Interaction in Breathy Voice, in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, I.R. Titze and R.C. Scherer, Eds., The Denver Center for the Performing Arts, Denver, CO, pp. 465-481 (1984).]

* * * * * *