Journal of Phonetics (1991) **19**, 425–432

Nonlinearities for one-dimensional propagation in the vocal tract

763

R. S. McGowan

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A.

Received 4th June 1990, and in revised form 16th January 1991

The possibility that nonlinearities can have a significant effect during sound propagation in the vocal tract is discussed. One-dimensional propagation is considered here, with only a mention of nonlinear instability of the glottal jet. A review of previous work suggests that a significant amount of energy exchange can take place between the harmonics of the voice due to nonlinear interaction. Acknowledging the existence of mechanisms for the nonlinear exchange of energy in the vocal tract in one-dimensional propagation does not alter the source–filter picture radically: the shape of the vocal tract still determines the resonance frequencies. The practical effect of these nonlinearities is to raise the question of what aspects of the speech spectrum should be attributed to the source and what to the filter.

1. Introduction

Clarification is needed in the relation between fluid mechanics and acoustics, because the sound we hear during speech production results from the unsteady motion of the fluid air in the vocal tract. We hear sound coming from the mouth because there is unsteady motion of air at the boundary between the mouth and the atmosphere. However, the fact that we hear sound coming from vocal tracts does not imply that the physics of air motion in vocal tracts is fully described by the equations of acoustic wave motion. These equations, known as the acoustic approximation, are linearized versions of the more general, nonlinear equations of motion. For articulatory synthesis, as commonly implemented, an equation for acoustic wave propagation in the vocal tract is solved using a numerical algorithm in either the time or the frequency domain. Therefore, to ensure accuracy in articulatory synthesis, it is necessary to examine the appropriateness of the acoustic approximation in speech production.

There are two questions to be asked. First, how much of the physics of air motion in the vocal tract during speech is captured by the acoustic approximation? In other words, is the acoustic approximation valid as a mathematical model of air flow in the vocal tract? Second, if there is physics not captured by the acoustic approximation, is this an important consideration in quantifying the sound that actually reaches our ears? It is possible that the acoustic mode is a small part of the air motion in the vocal tract and that the remainder of the unsteady air motion does not affect the sound that we hear. This paper addresses the first question, whether the physics of air flow in the vocal tract can be characterized by the acoustic approximation.

0095-4470/91/030425 + 08 $03.00/0

To justify the acoustic approximation in any situation, arguments are made based on a characteristic length scale, such as wavelength; a characteristic time scale, such as the period of oscillation; and characteristic magnitudes, such as the maximum pressure amplitude relative to ambient pressure. This is not a straightforward procedure, because these scales can vary with place and time. Consider wave propagation in a tube with a bend. In the straight parts of the tube with plane-wave propagation the only length scale is the wavelength (ignoring such things as viscous and thermal boundary layers). At the bend, the radius of curvature will enter as an important length scale, if it is small enough.

It has previously been argued that the air in the vocal tract during phonation has energy that is not associated with acoustic wave motion in the vocal tract (Teager & Teager, 1983; McGowan, 1988). This energy is associated with the rotational motion of the air that occurs when a jet exits from the glottis into the wider vocal tract. The jet is essentially a three-dimensional phenomenon: two dimensions for cross-sectional area, and one for direction of travel. It is possible that nonlinear interactions can give rise to instabilities of the jet, and this bears consideration. However, primary consideration will be given here to the nonlinearities resulting from one-dimensional volume flow in an unconstricted tube appropriate for vowel production, and behavior of the jet will receive only brief consideration at the end of this paper. Therefore, for the present purpose, the jet exiting from the glottis is assumed to fill the vocal tract tube immediately after exiting from the glottis.

2. Nonlinear steepening and damping

Two competing processes determine the importance of nonlinear terms in unsteady, one-dimensional air motion: In one, higher frequencies receive energy from lower frequencies, and in the other, energy is lost by dissipation mechanisms whose effects may increase with frequency.

Linear wave motion implies that energy propagates in each frequency separately and that energy is not exchanged between the frequencies. But, with nonlinear propagation, energy from the lower frequencies can be transferred to the higher frequencies. When energy undergoes such a transfer, the wave shape changes so that more abrupt pressure changes occur in space. For instance, for a compression wave, the minimum distance between a pressure maximum and pressure minimum becomes less as the wave travels further. It takes time for the process of energy exchange between the lower frequencies and the higher frequencies to occur, and as this occurs the distances between adjacent pressure maxima and pressure minima become smaller, and the wave is said to steepen. (Another way of describing the same phenomenon is to notice from the equations of motion that regions of high pressure travel at a higher speed than regions of low pressure, and this will have the effect of the pressure maxima catching up with neighboring minima in a travelling wave.)

The other competing process when considering the importance of nonlinearity in one-dimensional motion is that of energy dissipation or damping. Damping can be the result of phenomena that can be described by linear terms, such as viscous loss (see Keefe & Benade, 1983, for the effect of tube curvature on viscous loss), heat conduction losses and radiation losses, or it can be the result of nonlinear phenomena, such as the shedding of vorticity at a sharp corner (Ingard & Ising, 1967), acoustic streaming (Keefe, 1983) and sound–turbulence interaction (Ingard &

Singhal, 1974). The effect of damping is to decrease the amplitude of sound waves, and if the damping is greater at the higher frequencies, the wave will tend to become rounded. If the damping is of sufficient magnitude it will remove energy from any frequency that gained energy from nonlinear propagation. To assess the importance of nonlinearity in one-dimensional motion, it is necessary to compare the energy loss due to damping with the energy exchange due to nonlinear interaction.

3. The semi-infinite tube

It is instructive to consider the vocal tract to be a semi-infinite tube before moving on to the more realistic case of the finite length resonator. The semi-infinite tube illustrates the results of nonlinear steepening and damping.

One way to compare the importance of nonlinear steepening is to compare the sizes of terms in the equations of motion, based on an estimate of the time scales, length scales and the magnitudes of the dependent variables involved. It is a difficult task to express explicitly all the terms governing damping or energy loss in the vocal tract and to estimate their relative importance based on scaling arguments (cf. Liljencrants, 1985). Because of this difficulty, another, semi-empirical approach will be taken to estimate the magnitude of these loss mechanisms combined.

This alternative approach compares the time scales over which nonlinear interaction and damping have significant effect: the shorter the time scale the more important the effect. An analytical result can be used to estimate the time scale for nonlinear steepening. A time scale for damping can be estimated from spectral bandwidths. Recall that unconstricted tubes are being considered in this paper.

Suppose a piston is used to generate a positive pressure disturbance (compression wave) at one end of a semi-infinite tube, and that the frequency content of the disturbance is low enough for one-dimensional propagation to occur. (The higher order modes are neglected as more energy is pumped into the higher frequences.) If there are no dissipative mechanisms, then there will be a time when the minimum distance between a pressure maximum and a pressure minimum becomes zero, because of the steepening of the wave. This time is known as the onset time of shock, and it is when the wavefront becomes infinitely steep (Pierce, 1989). Using an approximate, second-order solution, the onset time of the shock is given as:

$$t = \frac{\rho c}{\beta \left| \frac{\partial p}{\partial x} \right|_{\substack{\text{initial} \\ \text{max}}}} \tag{1}$$

where ρ = density of air, c = adiabatic speed of sound, $\beta = 1.2$, and the partial derivative factor is the maximum pressure gradient at the initial time. The only variable in the above expression that is a property of the excitation, and not a property of the fluid alone, is the maximum initial pressure gradient.

The onset time of shock in a semi-infinite tube for a disturbance created by vocal fold closure will be estimated. (Although this is initially a rarefaction, it becomes a compression when reflected from the mouth. The "worst case" analysis is performed here, where the disturbance is considered to be a compression wave throughout its life.) Because equation (1) was derived using second-order approximations, it is consistent to use the first-order relation between pressure and particle velocity to estimate the pressure gradient in equation (1).

$$\rho \frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x} \tag{2}$$

where u = fluid particle velocity.

Using an early model of the glottal volume velocity pulse, the value of the maximum particle velocity time derivative can be estimated (Fant, 1979). With a maximum volume velocity of 500 cm^3/s, a steepness factor of 4 (which gives a 12 dB/octave roll-off in the voice spectrum), 125 Hz fundamental frequency, and a 4 cm^2 tube area, the maximum particle acceleration is 2.5×10^5 cm/s^2. This gives a shock onset time of 120 ms.

The effectiveness of damping mechanisms can be estimated by looking at the formant bandwidths. From Bogert (1953), the half-power bandwidths for the first three formants are between 67 Hz and 333 Hz. These bandwidths correspond to damping times (the time for the amplitude to decrease to 1/e of original value) of between 1 and 5 ms. It would appear, therefore, that dissipation mechanisms overwhelm nonlinear propagation mechanisms. However, the effects of nonlinear oscillation cannot be discounted in this worst case analysis, because significant amounts of energy can be exchanged between the harmonics on time scales of about one-tenth that of the onset of shock, or about 12 ms.

4. The resonator

The analysis above is instructive because it illustrates the physical mechanisms that must be incorporated when considering the importance of nonlinear phenomena in one-dimensional propagation. However, nonlinear phenomena can become important in driven resonators, even when they are unimportant in driven, semi-infinite tubes. Resonators are very effective at extracting energy from the driver at the frequencies of resonance, which means they are precisely the frequencies at which nonlinear phenomena are most likely to appear. While the order of magnitude of the amplitude for a travelling wave in a driven, semi-infinite tube is the same as that for the driver, in the finite-length resonator the amplitude of a standing wave near resonance can be much larger.

There has been some work on nonlinear phenomena in resonant tubes, open at one end and driven by a piston at the other (van Wijngaarden, 1968; Jimenez, 1973; Sturtevant, 1974; Keller, 1977, 1978). (It should be kept in mind that the results are for steady-state configurations and do not pertain to transience. In the case of travelling waves in semi-infinite tubes this simplification is not necessary to make things tractable.) For finite, driven tubes the factors deciding the relative importance of nonlinearity are the amplitude of the driver, the amount of damping, and the closeness of the driving frequency to resonance. As in the case of the semi-infinite tube, it is difficult to determine the loss mechanisms analytically, particularly in the case of the open ended tube where high amplitude radiation impedances are unknown.

One of the major differences between the driven resonator with open end and the driven semi-infinite tube, and even the driven resonator with a closed end, is that for the former case an initial compression will become a rarefaction upon reflection from the open end, then change back to a compression on the next reflection from the open end, and so on. In the latter cases a compression wave remains a compression wave throughout its life. The result of this is that while the nonlinear interactions steepening the wave in the semi-infinite tube and resonator with closed end are second order, the average nonlinear interactions for the open ended tube are third order (Jimenez, 1973). Among other things, this implies that for a given piston amplitude, the oscillation at resonance for the open tube will be larger than that for the closed tube, despite radiation losses. The quantitative implications of this result are discussed below.

Given that phonation into a neutral vocal tract configuration may be viewed as an open-ended tube driven by a high impedance flow source (piston), previous results on scaling can be used to estimate the importance of nonlinearities of the air in the vocal tract, neglecting jet effects. If δ is the ratio of the maximum piston velocity to the adiabatic speed of sound (peak Mach number of the piston) then the peak acoustic Mach number in the tube near resonance scales is $\varepsilon = \delta^{1/3}$ (Jimenez, 1973). (Peak acoustic Mach number is the maximum acoustic particle velocity divided by the speed of sound.) This scaling is a direct result of the fact that the nonlinearities are third-order for open-ended tubes. Assuming that the glottal jet fills the entire vocal tract, that the peak glottal Mach number is about 0.1, and that the ratio of vocal tract area to maximum glottal area is about 10, then $\delta = 0.01$. This gives $\varepsilon = \delta^{1/3} = 0.2$. This is a sizable Mach number, at which substantial nonlinearities have been observed by Sturtevant (1974) in open ended tubes. However, it should be noted that the amplitude of oscillation can be greatly diminished with such things as the addition of a plate with a hole cut into it clamped onto the opening (Sturtevant, 1974). The addition of the plate must either create a better impedance match between the tube and the atmosphere, resulting in more radiation loss, or more nonlinear resistance loss, or both.

Because the experimental conditions of Sturtevant may not reproduce all the loss mechanisms that are present in the vocal tract, it is helpful to go back to the theoretical results and, using an estimate of the overall energy loss, see if nonlinearity can be important. In his theory of nonlinearities of gas oscillations in open ended tubes, Jimenez (1973) models all losses by a parameter that expresses the deviation from perfect reflection at the open end. To second order in ε, the deviation from perfect reflection is given as $\varepsilon^2 b_2$, where b_2 is a dimensionless parameter. Further, nonlinear effects are most likely to result when the tube is driven near resonance. Deviations from resonance are given as $\varepsilon^2 \omega_2 f_0$, where f_0 is a resonance frequency according to linear theory, and ω_2 is a dimensionless parameter. The region of shock (infinite slope wavefronts) in the b_2–ω_2 plane has been plotted by Jimenez (1973), and is shown in Fig. 1.

To find a reasonable estimate of b_2, let $N =$ the number of round trips that a wave makes between the glottis and the lips before it is damped to 1/e of its initial amplitude. Then the expression $((1-(\varepsilon^2 b_2))^N = 1/e$ must hold, and once N is deduced, b_2 can be found. It has been previously estimated to take between 1 and 5 ms to damp a wave to 1/e of its initial amplitude in the vocal tract during vowel production. It takes about 1 ms for a wave to make a round trip from glottis to lips and back. Thus N is between 1 and 5. This gives b_2 between 4 and 16. From Fig. 1, shock formation does not occur for b_2 greater than 3. Further, there are other factors that could tend to lessen the importance of nonlinearity in these calculations. Recall that the peak Mach number of the glottal flow was assumed to be about 0.1.

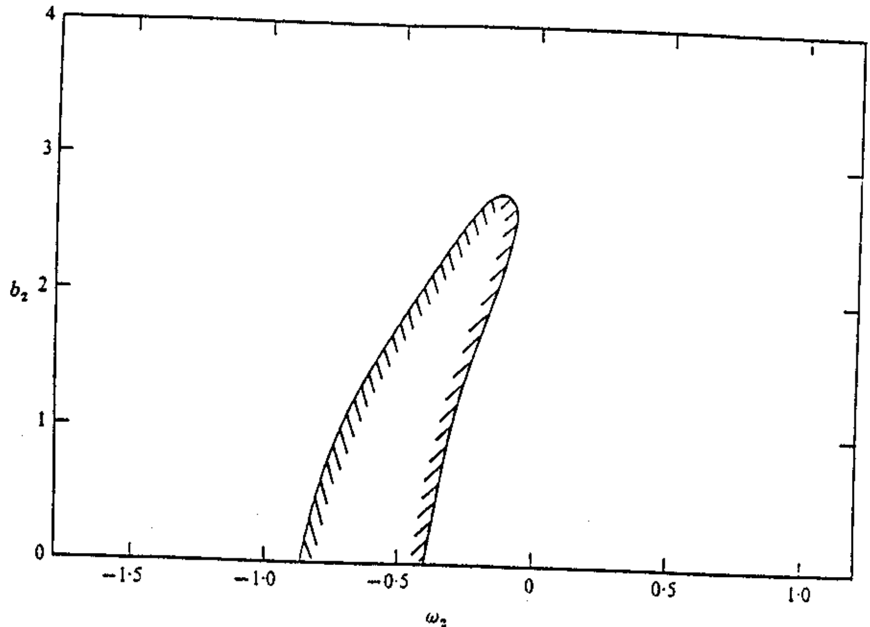

Figure 1. Region of shock production. Deviations from resonance frequency, f_0, are given by $\varepsilon^2\omega_2 f_0$, and deviations from perfect reflection are given by $\varepsilon^2 b_2$, where ε is a dimensionless amplitude parameter given in the text. ω_2 and b_2 are dimensionless and the shaded region indicates the extent of the region of shock. (From J. Jimenez, 1973, reprinted by permission of Cambridge University Press.)

But, this is the peak Mach number for the entire glottal flow, and any frequency components of the glottal flow near the frequency of the first formant may have a peak Mach number one or two orders of magnitude less than 0.1. This changes these rough calculations substantially. (A reduction in δ by a factor of 10 produces a reduction in ε^2 by about a factor of 0.2, thus increasing b_2 by a factor of 5.)

On the other hand, particularly for nearly coincident fundamental frequency and first formant frequency, the estimate of peak glottal Mach number near 0.1 may be the correct order of magnitude. Also, at this time, it cannot be stated how far away from the region of shock the system needs to be to neglect nonlinear terms, and with b_2 close to 4, the relative importance of nonlinearities should be considered.

If it turns out that nonlinearities are important to the physics of air motion in the vocal tract, why is it that they do not attract attention? A partial answer to this question may come with one more theoretical result. It is interesting to consider the perturbation from the resonance frequencies of linear theory for b_2 values smaller than 3. Assuming that $b_2 = 1.5$ and that $f_0 = 500$ Hz, the range of frequencies for which shock does occur is 488 to 495 Hz. If the region of strongest oscillation is considered to be resonance, then the resonance frequency has been perturbed by at most 12 Hz, even in the severe case of shock formation. Therefore, even in the case where energy is redistributed among the harmonics of the driver by nonlinear interactions strong enough to cause shock formation, the resonance frequency of the resonator remains relatively unperturbed.

5. Jet instabilities

Nonlinear interaction between the jet exiting the glottis and the rest of the air in the vocal tract has been excluded from the one-dimensional considerations here. However, it is possible that the jet remains thin while attaching to the walls of the vocal tract (Teager & Teager, 1983). This thin jet is a region of vorticity that can show instability with a delayed feedback mechanism. In experiments on jets with Mach numbers less than 0.15 flowing over cavities, the instability modes of the jet correspond to the acoustic modes of the cavity (East, 1966; Tam & Block, 1978). The glottal jet is within this Mach number range, so such an interaction in the vocal tract could give rise to enhanced vortex motion at frequencies that are acoustic modes of the vocal tract. While energy may be drawn from the acoustic field in the vocal tract this way, the air velocity at the lips may be enhanced at the acoustic resonance frequencies, thus enhancing the acoustic output at these frequencies. Also, the jet could strike surfaces in the vocal tract thus creating secondary acoustic sources. In both cases the frequencies closely corresponding to the resonance frequencies given by linear theory are the ones most greatly excited.

6. Conclusions

The nonlinear effects of the air motion cannot necessarily be ignored during vowel production, particularly in regard to the transfer of energy among the voice harmonics during intense voicing when the fundamental frequency is close to the first formant frequency. This conclusion follows from previous work on one-dimensional propagation in finite length, unconstricted tubes and an estimate of damping time of 5 ms or less. In the analysis of one-dimensional nonlinear effects, the shape of the vocal tract still determines resonance frequencies. Even in the severe case that the excitation is just strong enough to cause shock formation, which apparently does not occur in the vocal tract, the frequencies of resonance in finite tubes are still close to those given by linear theory. Three-dimensional effects of the glottal jet and its instabilities are not considered in these conclusions. However, a cursory review of the literature shows that in certain instances where three-dimensional effects are taken into account, nonlinear interaction between a jet and the rest of the air motion in a resonator may well enhance the response at the acoustic resonance frequencies. Thus, while, even in the nonlinear case, the vocal tract shape determines the resonance frequencies close to those given by the acoustic (linear) approximation, there may be a nonlinear exchange of energy from one frequency to another. This leads to questions about what spectral features are attributable to the filter and what to the source.

This work was supported by NIH grants DC-121 and HD-1994 to Haskins Laboratories. The author is grateful for advice from Patrick Nye and Philip Rubin.

References

Bogert, B. P. (1953) On the band width of vowel formants. *Journal of the Acoustical Society of America*, **25**(4), 791–792.

East, L. F. (1966) Aerodynamically induced resonance in rectangular cavities, *Journal of Sound and Vibration*, 3, 277–287.

Fant, G. (1979) Glottal source excitation analysis, *KTH Speech Transmission Laboratory QPSR*, No. 1, 85–107.

Ingard, K. U. & Ising, H. (1967) Acoustic nonlinearity of an orifice, *Journal of the Acoustical Society of America*, **42**(1), 6–17.
Ingard, K. U. & Singhal, V. K. (1974) Sound attenuation in turbulent pipe flow, *Journal of the Acoustical Society of America*, **55**(3), 535–538.
Jimenez, J. (1973) Nonlinear gas oscillations in pipes. Part 1. Theory, *Journal of Fluid Mechanics*, **59**, 23–46.
Keefe, D. H. (1983) Acoustical streaming, dimensional analysis of nonlinearities, and tone hole interactions in woodwinds, *Journal of the Acoustical Society of America*, **75**(5), 1804–1820.
Keefe, D. H. & Benade, A. H. (1983) Wave propagation in strongly curved ducts, *Journal of the Acoustical Society of America*, **74**(1), 320–1820.
Keller, J. J. (1977) Resonant oscillations in open tubes. *Journal of Applied Mathematics and Physics (ZAMP)*, **28**, 238–263.
Keller, J. J. (1978) Resonant oscillations in tubes with a mean flow, *Journal of Applied Mathematics and Physics (ZAMP)*, **29**, 85–91.
Liljencrants, J. (1985) Speech synthesis with a reflection-type line analog. Ph.D. dissertation KTH, Stockholm, Sweden.
McGowan, R. S. (1988) An aeroacoustic approach to phonation, *Journal of the Acoustical Society of America*, **83**(2), 696–704.
Pierce, A. D. (1989) *Acoustics*, pp. 569–571. Woodbury, NY: Acoustical Society of America.
Sturtevant, B. (1974) Nonlinear gas oscillations in pipes. Part 2. Experiment, *Journal of Fluid Mechanics*, **63**, 97–120.
Tam, C. K. W. & Block, P. J. W. (1978) On the tones and pressure oscillations induced by flow over rectangular cavities, *Journal of Fluid Mechanics*, **89**, 373–399.
Teager, H. M. & Teager, S. M. (1983) Active fluid dynamic voice production models, or there is a unicorn in the garden. In *Vocal Fold Physiology* (I. R. Titze & R. C. Scherer, editors), pp. 387–401. Denver: The Denver Center for Performing Arts.
Wijngaarden, L. van (1968) On the oscillations near and at resonance in open pipes, *Journal of Engineering Mathematics*, **2**, 225–240.