

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



The use of low-frequency ultrasonics in speech processing

Farzaneh Ahmadi and Ian McLoughlin
*Nanyang Technological University
Singapore*

1. Introduction

Audible sound analysis is assumed to be an integral part of any speech processing system, since the audible frequency ranges are naturally used for vocal communications. Inaudible sound, either ultrasonic or subsonic, is less widely researched – although sub-sonics (also called infrasound) have found a niche application in sound strengthening for spatial effects (Begault, 1994) and more immersive audio experiences (Kyriakakis, 1998) since there is evidence that these lower frequencies can be felt even if not heard. Higher frequency ultrasound is commonly used for diagnostic imaging in both medical and engineering fields (Szabo, 2004), and even sometimes for medical treatment (Haar, 1999). Due to the required resolution of these applications, they tend to operate at frequencies above the MHz range. In the animal kingdom, the majority of communication signals share a similar frequency range with humankind, although extending downward to infrasound in sea creatures such as whales (Clark, 2004), and upward to ultrasound in bats, moths, dolphins and so on (McLoughlin, 2009).

One phonetically significant and growing application of ultrasound is ultrasonic speech, as a contribution of ultrasound to speech analysis and processing. The technology acts to augment the human natural speech production system, using data extracted from ultrasonic analysis to synthesize elements of audible speech. In this application, an ultrasonic signal in the kHz range is injected into the vocal tract (VT), which propagates and resonates through the vocal cavities and is emitted from the mouth as ultrasonic speech. This signal could potentially be converted to audible speech by appropriate down-conversion.

This chapter introduces ultrasonics, particularly low frequency (LF) ultrasonic waves, and analyses their interaction with the human vocal tract. Ultrasonic frequencies used in this application are relatively low (near-audible), extending upwards from the upper threshold of human hearing to around 100 kHz in frequency. These frequencies are easily generated, captured and recorded by much existing audio equipment – and can be readily processed in real-time with available hardware.

In this chapter, the basic principles of ultrasonic speech technology and the most important issues concerning its implementation and signal processing are addressed. Most importantly, a modelling scheme describing ultrasonic behaviour within the VT is proposed and proven as the theoretical basis and the main framework of implementation of ultrasonic speech. This chapter concludes with open areas of research concerning the technology.

The scope of this chapter is as follows: In order to have a precise understanding of the problem, first the attributes of ultrasonic propagation are analyzed physically and mathematically in section 2. This section investigates these attributes, and describes linearity preconditions of any gas medium, the compliance with which, would allow ultrasonic propagation in that medium to be considered linear and lossless.

Section 3 analyses the plausibility of the linearity assumption for the propagation of the low frequency portion of the ultrasound bandwidth in the VT by a numerical analysis of the impact of dispersion and attenuation of LF ultrasound and addresses issues such as exhaled CO₂ as a dispersive wave medium for ultrasound, losses and cross modes of resonance of the VT in such frequencies.

Given this basic perspective, section 4 introduces ultrasonic speech as the usage of LF ultrasound for speech processing, surveys previous implementations of the technology and describes the necessary requirements of the implementation. As in this method, the human VT is used to produce the ultrasonic output signal, there is a need to study the anatomy and physiology of human speech production system in general in section 5. The necessary preconditions for linear modelling in section 2 along with the numerical analysis of section 3, lead to the derivation of a linear source-filter model for the ultrasonic speech process in section 6. Many applications in the theory of speech processing rely on the classical source-filter model of speech production. Section 6 considers how this model can be adapted to ultrasonic wave propagation in the vocal tract by manipulating the sonic wave equations and deriving the vocal tract transfer function for ultrasonic propagation.

At audible frequencies, linear predictive analysis (LPA) applies a linear source-filter model to speech production, to yield accurate estimates of speech parameters. Section 7 investigates the possibility of extension of LPA to cover ultrasonic speech. Discussing some simplifying assumptions, the section leads to the application of LPA for the analysis of ultrasonic speech. By the extension of LPA to ultrasonic speech, we introduce the main set of features needed to be extracted from the ultrasonic output of the VT to be utilized in speech augmentation. The chapter then presents a concise outline of current research questions related to this topic in section 8. Section 9 finally concludes the discussion.

2. Attributes of ultrasonic propagation

Ultrasound can be defined as “Sound waves or vibrations with frequencies greater than those audible to the human ear, or greater than 20,000 Hz” (Simpson & Weiner, 1989). The starting point of the ultrasonic bandwidth resides implicitly somewhere between 16-20 kHz due to variations in the hearing thresholds of different people. The bandwidth continues up to higher levels¹ where it goes over to what is conventionally called the hypersonic regime (David & Cheeke, 2002). The upper limit of ultrasound bandwidth in a gas is around 1 GHz and in a solid is around 10¹³ Hz (Ingard, 2008). At such mechanical vibrations exceeding the GHz range, electromagnetic waves may be emitted so that the upper limit of ultrasound may induce RF (radio frequency) electromagnetic waves (Lempriere, 2002).

The general definition of sound indicates that “sound is a pressure-wave which transports mechanical energy in a material medium” (Webster, 1986). This definition can extend the

¹ which in a gas is of the order of the intermolecular collision frequency and in a solid is the upper vibration frequency (Ingard, 2008).

margins of understanding of sound beyond the hearing limitations of humans to cover any pressure wave including ultrasound. It has to be noted that similar to the sense of sight, which subjects the visible light region of the EM spectrum to special attention, the human sense of hearing has differentiated the “audio” segment of sound to be classically termed as “sound” in common language and other portions of the bandwidth have thus been classified in relation to the audible part as ultra or infrasound (similarly to visible light and infrared, ultraviolet terminology).

The fact which should not be concealed is that the audible sub-band is only a tiny slice of the total available bandwidth of sound waves, and the full bandwidth, except at its extreme limits can be described by a complete and unique theory of sound wave propagation in acoustics (David & Cheeke, 2002). Accordingly all of the phenomena occurring in the ultrasonic range occur throughout the full acoustic spectrum and there is no propagation theory that works only for ultrasound.

The theory of sound wave propagation in certain cases simplifies to the theory of linear acoustics which eases linear modelling of acoustic systems. It is generally preferential to approximate a system with a linear model where the assumptions of such modelling are plausible. Ultrasound inherits some of its behaviours from its nature of being a sound wave. There are also characteristics of the medium which impose some medium specific constraints on ultrasonic waves. Based on these facts we will review the general characteristics of ultrasound propagation as a sound wave and the effects of the medium, paying special attention to the required pre-conditions of linearity.

2.1 Wave based attributes of sound

Ultrasound as a sound wave, obeys the general principles of wave phenomena. The theory of wave propagation stems from a rich mathematical foundation of partial differential equations which are valid for all types of waves (Ikawa, 2000). In other words every wave, regardless of its production and physical detail of propagation can be described by a set of partial differential equations. All common behaviours observed in waves are mathematically proven by these equations (Rauch, 2008).

To rest under the scope of generalization of the theory of waves, a physical phenomenon solely needs to fulfil the preconditions of being a wave by complying with the restrictions imposed by the wave equations. Afterwards the common behaviour of waves, proven mathematically for the solutions of these equations, would be valid for that specific physical phenomenon too. It has to be noted that although in today’s understanding of waves we are quite confident that for example, sound “is” a wave, however compliance of each wave type with the wave equations as the necessary pre-condition, has long ago been proven by scientists of the corresponding discipline (Pujol, 2003).

When the dimensions of the material are large in comparison to the wavelength, the wave equations become further simplified and can approximate the wave propagation as rays². These simplified sets of wave equations are the basis of geometric wave theory (aka ray theory) of wave propagation (Bühler, 2006). The geometric wave theory permits freedom of microscopic details of wave propagation and describes the wave movement, reflection and refraction in terms of rays. The theory has been initially observed in optics and owes its

2 A ray is a straight or curved line which follows the normal to the wave-front and represents the two or three dimensional path of the wave (Lempriere, 2002).

application to acoustic waves to (Karal & Keller, 1959; 1964) and has yielded geometric acoustics (Crocker, 1998) as the dual to wave acoustics (Watkinson, 1998).

As a high frequency approximation solution to the wave equations, ray theory fails to describe the wave phenomenon in low frequencies when the wavelength is large compared to the dimensions of the medium. Consequently, in low frequencies we have to refer to general wave equations as the wave theory to describe the wave phenomenon. It has to be noted that wave theory is always valid but only in smaller wavelengths in comparison to the dimensions of the medium can the analysis be simplified by the geometric theory.

In any case, because all the waves obey the same sets of partial differential equations, they have common attributes which are guaranteed by several principles extracted out of the wave equations. These principles manifest geometric and wave behaviour and are the general laws which impose similar conditions upon the propagation of waves in microscopic and macroscopic scales. The Doppler effect (Harris & Benenson et al., 2002), principle of superposition of waves in linear media (Avallone & Baumeister et al., 2006), Fermat's (Blitz, 1967) and Huygens principles (Harris & Benenson et al., 2002) are the fundamental laws of propagation for all the waves including ultrasound in wave and geometric theory. For interested readers, the mathematical derivation of some of these principles using wave equations is covered in (Rauch, 2008).

For universal wave events such as diffraction, reflection and refraction which obey the general principles of wave propagation, there would be no exception to the general theory of sound propagation for ultrasound (David & Cheeke, 2002) except only the change of length scale which means that we have moved to different scales of the wavelength so the scale of material in interaction with waves and the technologies used for generation and reception of these waves will be different (David & Cheeke, 2002).

2.2 Medium based attributes of sound

The exclusive wavelength-dependant behaviours of ultrasound will present itself in the influence of the medium on wave propagation and we expect to observe some differences with audible sound where the wave propagation is apt to be influenced by characteristics of the medium through which it travels. In this section we consider the general attributes of a medium which impose special behaviours on a sound wave. Next in section 2.3 we will consider the effect of such attributes on ultrasound waves. When the medium of sound wave propagation is considered, the first important attribute under question is the linearity of the medium. Also important is a consideration of the attenuation mechanisms by which the energy of a sound wave is dissipated in the medium.

2.2.1 Linearity

Propagation of sound involves variations of components of stress (pressure) and strain in a medium. For an isolated segment of the medium we may consider the incoming wave stress as the input and the resulting medium strain as the response of the system to that input. To consider a medium of sound propagation as a linear system the stress-strain relation should be a linear function around the equilibrium state (Sadd, 2005). Gas mediums such as the air, match closely to the ideal gas law in their equilibrium state (Fahy, 2001) which states that:

$$\tilde{p}\tilde{v} = nR\tilde{T} \quad (1)$$

Where \tilde{p} is the gas pressure, \tilde{v} is the volume, \tilde{T} is temperature and n , R are constant coefficients depending on the gas. If one of the three variables of \tilde{p} , \tilde{v} or \tilde{T} remains constant, the relation of the other two, can easily be understood from (1) but sound wave propagation generally alters all of these three components in different regions of the gas medium. A general trend is to consider sound wave propagation in an ideal gas as an adiabatic process meaning no energy is transferred by heat between the medium and its surroundings when the wave propagates in the medium (Serway & Jewett, 2006). If the ideal gas is in an adiabatic condition we would have (2) as the relation of pressure (\tilde{p}) and density ($\tilde{\rho}$) where α is a constant and the exponent γ is the ratio of specific heats at constant pressure and constant volume for the gas (which has the value 1.4 for air) (Fahy, 2001):

$$\tilde{p} = \alpha \tilde{\rho}^\gamma \quad \rightarrow \quad \partial \tilde{p} / \partial \tilde{\rho} = \gamma (\tilde{p} / \tilde{\rho}) \quad (2)$$

Equation (2) does not generally demonstrate a linear relation between pressure and density in an ideal gas but in small variations of pressure and density around the equilibrium state, $\tilde{p}/\tilde{\rho}$ can be considered to be constant and we will have:

$$[\partial \tilde{p} / \partial \tilde{\rho}]_0 = \gamma \left(\frac{p_0}{\rho_0} \right) \rightarrow [\partial \tilde{p} / \partial \tilde{\rho}]_0 = K / \rho_0 \quad (3)$$

where $[\partial \tilde{p} / \partial \tilde{\rho}]_0$ denotes small variations around the equilibrium, p_0 and ρ_0 are the pressure and density of the gas at equilibrium and constant $K = \gamma p_0$ is called the adiabatic bulk modulus of the gas (Fahy, 2001). Based on the above discussion the linear stress-strain relation in an ideal gas medium can be considered to exist between variations of pressure ($\partial \tilde{p}$) and variations of density ($\partial \tilde{\rho}$), having an adiabatic process (no loss) and small variations of pressure and density around the equilibrium.

2.2.2 Dissipation mechanisms

In section 2.2.1 we observed that under three conditions of having an ideal gas with an adiabatic process (no loss) and small variations of pressure and density around the equilibrium as a result of sound wave, air can be considered a linear lossless medium of sound wave propagation. These assumptions are known to be reasonable for audible sound but we need to consider their validation for the ultrasound case. Although we can preserve the small pressure variations precondition of linearity for ultrasonic speech application, as we will observe shortly, the physics of the problem make the assumptions of an adiabatic process and ideal gas behaviour of the air for ultrasonic frequencies, to be more of an approximation.

We need to consider the effects of this approximation i.e. attenuation (heat loss) and also deviation of the air from linear state equation (3) of an ideal gas in the frequency range of LF ultrasound. These derivations could cause dissipative behaviours in the air medium of sound propagation as a result of several phenomena including viscosity, heat conduction and relaxation. We will describe each briefly.

2.2.2.1 Viscosity and heat conduction

Viscosity is a material property that measures a fluids resistance to deformation. Heat conduction on the other hand is the flow of thermal energy through a substance from a higher to a lower-temperature region (Licker, 2002). For air, viscosity and heat conduction are known to have negligible dispersive effects (section 2.3.4) for sound frequencies below

50 MHz (Blackstock, 2000) but these mechanisms cause absorption of sound energy. Their effect in an unbounded medium can be considered by introducing a visco-thermal absorption coefficient α_{tv} to the time harmonic solution of the wave equation, the amount of which demonstrates the necessity of switching to wave equations in thermo-viscous fluids for the analysis of waves in frequency range of interest.

2.2.2.2 Relaxation

Gases demonstrate a behaviour called relaxation in sound wave propagation. Relaxation denotes that there is a time-lag (relaxation delay time) between the initiation of the disturbance by the wave and application of this disturbance to the gas which is compared to the time a capacitor needs to reach its final voltage value in an RC circuit (Ensminger, 1988). This delay could result from several physical phenomena. First the viscosity, second heat conduction in the gas from the places which the wave has compressed to the places where the wave has rarefacted which will cause the energy of the wave to be distributed in an unwanted pattern delaying the energy from returning to the equilibrium. The third and the most important case of relaxation in LF ultrasound applications is the molecular relaxation resulting from the delays of multi-atomic gas molecules having several modes of movement, vibration and rotation and the delay for molecules to be excited in their special vibration mode (Crocker, 1998).

When a new cycle of the wave is applied to the relaxing medium, the delay between the previous cycle of the wave disturbance and the resulting response of the medium will consume some of the energy of the new cycle, to return the medium to its equilibrium. This will cause absorption of the wave energy which depends on the frequency of the wave and the amount of the delay. In addition, due to the relative variations of frequency and relaxation delay, waves of some frequency can propagate faster than other frequencies. Consequently, relaxation in the gases is the physical cause of frequency dependant energy absorption and dispersion of the wave. As for this being a reason for dispersion, readers may refer to a mathematical discussion in (Bauer, 1965), while for the absorption as a result of relaxation, the interesting discussions in (Ingard, 2008) and (Blitz, 1967) should be consulted.

2.3 Effects of the medium on ultrasound propagation

Having considered the dispersive mechanisms of a gas for ultrasound frequencies, now we can consider the effects of these mechanisms in attenuation and dispersion of ultrasound. We will also discuss the case of resonance in the medium of ultrasonic propagation because these analyses will finally be applied to the propagation of ultrasound in the vocal tract which is a resonant cavity.

2.3.1 Speed

The sound speed in a medium (not necessary linear) has been formulated by (Fahy, 2001) as:

$$c^2 = \partial \tilde{p} / \partial \tilde{\rho} \quad (4)$$

While a gas medium maintains a linear behaviour as an ideal gas, based on the discussion of section 2.2.1, this speed is not a function of frequency and is evaluated according to the formula (Blackstock, 2000):

$$c = \sqrt{K/\rho_0} \quad (5)$$

If the phase speed of sound propagation in a medium is independent of the frequency as per (5), the medium is non-dispersive (Harris & Benenson et al., 2002), and all the events which rely on the speed of propagation (such as refraction) will be similar for sound waves across the whole frequency range (including ultrasound and audio) in that medium.

2.3.2 Acoustic impedance

The concept of acoustic impedance³ is analogous to electrical impedance and is defined as the ratio of acoustic pressure \tilde{p} and the resultant particle velocity \tilde{u} (Harris & Benenson et al., 2002). Impedances determine the reflection and refraction of waves over medium boundaries. In a homogenous material the acoustic impedance is a material characteristic, so it is called characteristic acoustic impedance and is formulated as:

$$Z = \frac{\tilde{p}}{\tilde{u}} = \rho_0 \cdot c \quad (6)$$

Where ρ_0 is the density of undisturbed medium and c is the speed of sound (The formula is same for both solids and fluids when they are homogenous). From (6) it is observed that in a non-dispersive material the acoustic impedance is independent of the frequency, so the impedance based characteristics (such as reflection coefficients) will be general to the case of all sounds in a non-dispersive medium (Harris & Benenson et al., 2002).

2.3.3 Attenuation

Attenuation is the loss of the energy of sound beam passing through a material. Attenuation can be the result of scattering, diffraction or absorption (Subramanian, 2006). Scattering and diffraction losses are not of much concern in the current application of LF ultrasounds in the vocal tract so we are going to discuss absorption in more detail.

The main causes of absorption of energy in gases in ultrasound frequencies are the molecular relaxation and visco-thermal effects. Visco-thermal effects introduce a visco-thermal absorption coefficient α_{tv} while molecular relaxation introduces several molecular coefficients α_{M_i} for each of the M_i gases in an N gas mixture (like air). The total absorption coefficient α is the sum of these values (Blackstock, 2000).

$$\alpha = \alpha_{tv} + \sum_{i=1}^N \alpha_{M_i} \quad (7)$$

α_{tv} is a scalar multiplicand of f^2 , (f being the frequency of the sound wave) while α_{M_i} is a scalar multiplicand of $\frac{f^2}{f^2 + f_r^2}$ (f_r is the relaxation frequency of the gas⁴) (Blackstock, 2000).

The impact of absorption is usually regarded by the value of absorption coefficient. In an unbounded medium for the time harmonic analysis of the wave, the role of absorption coefficient α would be an exponential multiplicand $e^{-\alpha r}$ to be multiplied by the lossless wave solution where r is the distance of the inspection point from the source. In bounded

3 The unit for acoustic impedance is $\text{Kg/m}^2/\text{s}$ and is called Rayl, named after Lord Rayleigh.

4 $f_r = \frac{1}{2\pi\tau}$ where τ is the relaxation time delay of the gas.

media we need to switch to damped wave equations to consider the effect of absorption. Absorption is usually accompanied by dispersion (Blackstock, 2000).

2.3.4 Dispersion

There are several possible causes for dispersion in a gaseous medium among which viscosity, heat conduction and relaxation are the most applicable for propagation of ultrasound frequencies. It is known that the dispersive effects of viscosity and heat conduction in air at frequencies below 50 MHz are negligible (Blackstock, 2000), so the main cause of dispersion in lower frequency ultrasound will be molecular relaxation (Blackstock, 2000). Sound speed in a relaxing gas with standard temperature and pressure is computed by (Crocker, 1998):

$$\frac{c^2}{c_0^2} = 1 + \frac{\epsilon}{1 + \epsilon} \cdot \frac{\omega^2 \tau^2}{1 + \omega^2 \tau^2} \quad (8)$$

c is the speed at angular frequency $\omega = 2\pi f$, ϵ is the relaxation strength and τ is relaxation time which are constants for a specific gas. c_0 is the low frequency speed of sound in the gas. The value $\omega\tau = 1$ occurs at the relaxation frequency f_r and the effect of dispersion in frequencies around f_r is more intense. For example CO_2 introduces dispersion at ultrasonic frequencies around 28 kHz (Dean, 1979).

2.3.5 Resonance

An important attribute of some sound propagation media is resonance at certain frequencies. Resonance is tied closely with the presence of standing waves in a medium. A resonant medium for sound waves should first have the possibility of forming standing waves and second the capability of frequency selectivity. Standing waves are normally formed as a result of interference between two waves travelling in opposite directions. For an interesting description of how standing waves are formed in an open-closed end tube as a simplified model of vocal tract, readers may refer to (Johnson, 2003).

The major cause of resonance for sound waves of certain frequencies in a medium is the geometric structure of that medium. When the geometry is more suitable for sound waves of certain frequencies to be distributed as standing waves in the medium e.g. the medium dimensions are wider where the standing wave has a rarefaction and narrower where it has a compression point, resonance can happen at that frequency. The resonance frequencies of an open/open and closed/open tube are a clear example of this (Halliday & Resnick et al., 2004). For the case of interest, namely ultrasonic propagation through the vocal tract, we need to emphasize that the resonant behaviour of the VT will have one major difference with the audible case. In audible frequencies, due to the relatively large wavelength of the sound, standing wave patterns establish mainly along the axial length of the tract. But as we move toward lower wavelengths, in addition to axial standing waves, cross-modes of resonance can be created across the width of the tract, resulting in more complex patterns of resonance. Analysis of these cross-modes urges us to consider three dimensional equations for ultrasonic wave propagation in the tract while in audible range we normally consider the one dimensional wave equation.

Now that we have understood the main characteristics of ultrasound and its deviations from the general sound category in terms of attenuation and dispersion, we will consider a numerical analysis of the impact of these characteristics in LF ultrasound.

3. Low-frequency ultrasound

A major application of ultrasound is scanning, both in medical and industrial applications, relying upon reflections of the wave by an object (such as a defect in non destructive testing or a human fetus in ultra-sonography). When the dimensions of the reflecting object are smaller than the wavelength, the wave does not reflect back but scatters as an unfavourable wave behaviour. So to detect a defect, one needs to use a wavelength equal or smaller than its dimensions e.g. for a defect size of millimetres we need to use a sound wave above MHz frequency (Subramanian, 2006). The demand for detecting smaller details moves us out of audible range to use higher ultrasound frequencies, limiting the application of LF ultrasound to special cases such as cavitation or industrial non destructive testing.

Low Frequency ultrasound in ultrasonic speech application is considered as a portion of the ultrasonic bandwidth, starting from human hearing threshold up to 100 kHz. We will discuss the reasons for selection of this portion of the bandwidth shortly. As we will see in this section, LF ultrasound has properties which make it a suitable substitute for audible excitation of the vocal tract to produce ultrasonic speech.

The discussion of this section is biased so that the numerical analysis will provide us with an insight about the impact of attenuation and dispersion effects of LF ultrasound propagation in the vocal tract which we should discuss before being capable of modelling ultrasonic speech process as a linear and lossless system.

We are going to consider attributes of LF ultrasonic propagation in the air, and through the air-tissue interface. Soft body tissues and the air in the vocal tract are the regions of interest for ultrasonic speech production and both can be considered as homogeneous fluids (Zangzebki, 1996). Sound waves in the volumes of fluids are longitudinal (Fahy, 2001) so the mode of ultrasound propagation in the vocal tract and soft tissues of our concern will be longitudinal. As we will see in this section, high reflection coefficients of the air-tissue interface will reflect back most of the ultrasound wave energy over vocal tract walls, so we do not need to consider LF propagation through human body tissue.

3.1 Propagation through air-tissue interface

As described in (Caruthers, 1977), if the wavelength of the wave is small enough in comparison to the dimensions of the boundary of two media, Fermat principle will govern and the wave will be reflected with an angle (to the normal) equal to the angle of incidence. The reflection coefficient (Crocker, 1998) determines the proportion of energy to be reflected. Referring to (Zangzebki, 1996), we observe that the acoustic impedance of the air is too small in comparison to other materials of our problem. The reflection coefficient for an air-tissue interface (acoustic impedance $Z_1=0.0004 \times 10^6$ Rayls for air and $Z_2=1.71 \times 10^6$ for muscle)⁵, is computed to be -0.99 (same value with positive sign for the tissue-air interface)⁶.

⁵ Speed of sound is approximated 1600 m/s in muscle and 330 m/s in the air.

⁶ The minus value merely indicates the phase difference between the incident and reflected signal to be 180 degrees.

The value illustrates that ultrasound will almost completely reflect back from an air/tissue or tissue/air interface. This is expected also by the impedance mismatch effect (Zangzebski, 1996).

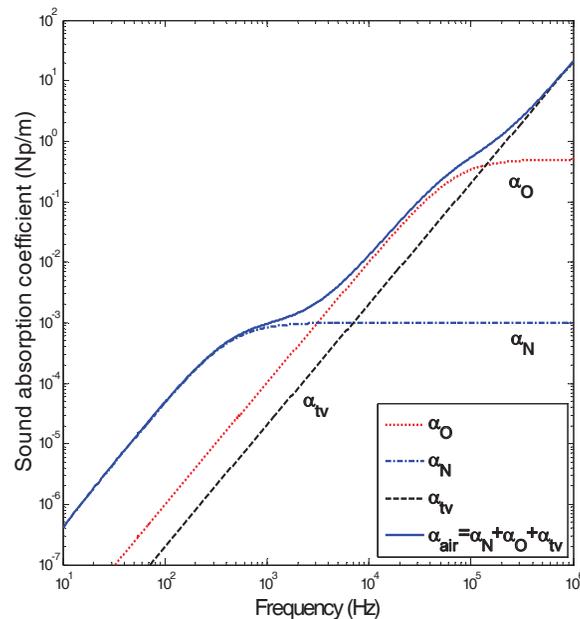


Fig. 1. Variation of the absorption coefficient of the air with frequency

3.2 Propagation through the air

In ultrasonic speech applications, the ultrasonic signal entering the vocal tract from the transducer has to travel through the air bounded by VT walls. As the exclusive effects of the medium on ultrasound, attenuation and dispersion are frequency-dependant, we need to have a numerical overview of the significance of these effects on ultrasound propagation in the air.

3.2.1 Attenuation

The absorption coefficient α was introduced in section 2.3.3 to be a sum of visco-thermal α_{tv} and molecular relaxation coefficients. For the air the two major components of oxygen and nitrogen have the molecular relaxation coefficients of α_N and α_O . Figure 1 demonstrates the variation of value of α (being equal to $\alpha_{tv} + \alpha_N + \alpha_O$) with frequency. As the figure demonstrates, this value reaches around $0.1 N_p/m$ in sound frequency of 100 KHz which is less than 1 dB/m.

3.2.2 Dispersion

As stated in 2.2.1 and 2.3.1, one precondition of linearity for ultrasound propagation in air is that the air medium should be an ideal gas in which the speed of sound is independent of sound frequency. For frequencies in the ultrasonic range, air deviates from this attribute as a result of being composed of dispersive carbon dioxide (CO_2) which should be considered in the VT due to the higher proportion of CO_2 in the exhaled air flow (The percentage of CO_2 in exhaled air is 4% which is 100 times that in normal air (Zemlin, 1997). This deviation initiates at frequencies above 28 kHz (Dean, 1979) and needs to be addressed here in detail.

The visco-thermal dispersion of sound in air for frequencies below several hundred MHz, depends on the square of the frequency but is negligible for frequencies between 1 Hz and 50 MHz at STP⁷ (Blackstock, 2000; Dean, 1979). Thus there remains only molecular relaxation dispersion. Among the main components of air (nitrogen, oxygen, carbon dioxide and water), nitrogen and oxygen can be considered non-dispersive as the maximum variation of sound speed in these two gases with the increase of frequency from zero to infinity is only a few centimetres per second (Blackstock, 2000). Water and carbon dioxide have effects on variation of sound speed with frequency in the air. Specifically, pure carbon dioxide in which the speed of sound may vary about 8m/s between frequencies of 1kHz and 100 kHz (Crocker, 1998).

Equation (8) demonstrated the dispersion characteristics of the gas, and is shown in figure 2. The same figure is reported for air, which illustrates that the dispersive effect of humid air is negligible for frequencies up to 5 MHz (Crocker, 1998).

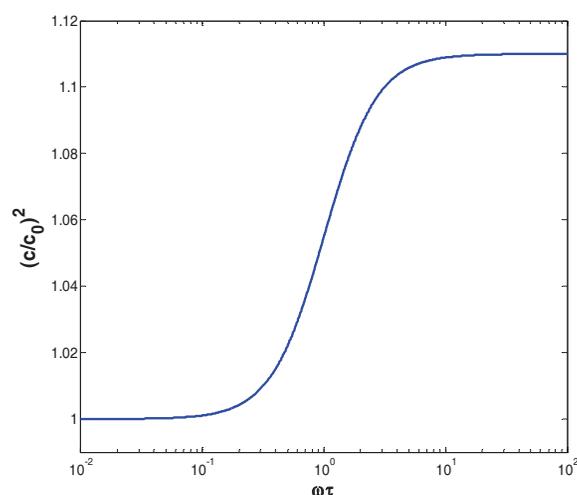


Fig. 2. Dispersion characteristics of a relaxing gas mixture

Based on studies of sound propagation in the atmosphere (Dean, 1979), the resulting variation of sound speed in air as a mixture of these gases (which obeys figure 2) over frequencies up to 5 MHz is in the order of few cm/s (for sound speed of approximately 343 m/s at STP). Referring to the monotonic pattern of increase of sound speed in (8) and figure 2, where the maximum speed variation for air at frequencies up to 5 MHz is negligible, and considering the percentage of gases other than carbon dioxide in the air, the dispersive effects of air can confidently be considered negligible for the dimensions of the vocal tract and the frequency range of interest (namely, less than 100 kHz).

As a conclusion of the preceding discussion, for ultrasonic frequencies of less than 100 kHz, and for the dimensions of our problem the air only has the effect of frequency dependant attenuation with an absorption coefficient of less than 1 dB/m and can be considered as a lossless non-dispersive linear medium in modelling ultrasonic propagation in the vocal tract. Linear systems are considered preferential for speech analysis and processing, and so we would prefer to limit our application to frequency ranges which can assure a linear relationship, if possible.

⁷ Standard temperature and pressure.

4. Application of LF ultrasound in speech augmentation

Having described the preliminary basics, we now turn our attention to the application of ultrasound in speech augmentation. We will divide these applications into two sets. The first set corresponds to applications in which ultrasonic excitation can act as a substitute to replace the natural excitation of the human voice production system. In this case, a person can speak without any voicing and an ultrasound to audible conversion system can produce a final audible sound. In the second set, ultrasonic excitation can be considered to act as a supplement to the natural excitation to provide additional data from the vocal tract for computational analysis.

Examples of the former set apply to people who suffer from impairments to their voice box and are incapable of producing natural excitations in their VT including laryngectomised patients and the voice-rest cases (Pozo, 2004). Another example is where audible speech is highly affected by surrounding or background noise and common levels of conversation or even high amplitude speech cannot be heard, such as at airports, on the battlefield, or in industrial environments (MacLeod, 1987). The other application in this set is when one does not wish to be heard in cases of talking in private places or when being heard will disturb other applications of a system like dictation in human-computer interfaces of crowded offices.

For the examples of the second set we may primarily consider ultrasound for providing additional data in speech recognition systems aiming to achieve higher levels of robustness. As another application in this set, we can mention cases where ultrasound can be augmented as an auxiliary excitation to the VT to provide voicing information when converting whispered speech to normally phonated speech. In this application, while a person whispers, the unvoiced segments of speech are extracted from the whispered signal but the voiced segments are reconstructed using the VT resonance data extracted from the ultrasonic output of the VT. This special augmentation can be used in whispered speech communications over telephone, and speech aids for people who have to speak in whisper mode for medical reasons.

4.1 Ultrasonic speech

In this chapter the application of LF ultrasonic waves in speech augmentation is termed ultrasonic speech. By ultrasonic speech we mean a system which augments an ultrasonic excitation to the human voice production mechanism as a substitute or supplement to the natural excitation and extracts feature sets from the resulting ultrasonic output to be used in several tasks including conversion to the audible speech, speech regeneration, recognition, enhancement and communication. The signal which is injected from an ultrasonic transducer to the VT via several possible injection points propagates through the tract and emits out of the mouth, where it is picked by another transducer and is delivered to the processing algorithms in charge of feature extractions in the ultrasonic domain or the equivalent audible domain. The set of these extracted features are then delivered as the output of the ultrasonic speech system to other modules which may pursue classic tasks of speech generation, recognition, and so on.

The ultrasonic frequency range of this application starts from the higher threshold of human hearing up to around 100 kHz. As stated before, this frequency range has some characteristics which suit the propagation of ultrasonic waves in the vocal tract to be

modelled in linear and lossless acoustic domains. In this domain we can be equipped with facilities of linear modelling of the VT behaviour in response to ultrasonic excitation.

4.2 Previous implementations

Speech processing science relies heavily on data provided by ultrasonic scanning of the position of VT articulators as an indirect contribution of ultrasound to speech processing (Kelsey & Minifie et al., 1969). As an example we can mention the data provided by real-time ultrasonic monitoring of the tongue (Shawker & Sonies, 2005) to speech processing. In direct applications, ultrasonic waves are used directly to produce an ultrasonic speech signal which is sought for speech processing features (MacLeod, 1987). Similarly, an audible signal modulated by an ultrasonic carrier in ultrasonic communication (Akerman & Ayers et al., 1994), or converted to audible speech as a consequence of the non-linearities of the system in ultrasonic hearing (Lenhardt & Skellett et al., 1991).

These are niche examples of several contributions of ultrasonics to speech processing, yet there are few examples of the implementation of low frequency ultrasound in speech augmentation (ultrasonic speech). To consider further, let us first review the implementations of these methods.

The history of ultrasonic speech goes as far back as 1987 when MacLeod filed a patent for a non audible speech generator system (MacLeod, 1987). The system augmented a series of pulses similar to the glottal pulse shape in ultrasonic frequency range of 15 to 105 kHz to the vocal tract. MacLeod considered the output at the mouth as being an amplitude modulation of the ultrasonic input. He then proposed the idea of passing the output to an ultrasonic detector where it was down converted to audible range to pursue a further goal of synthesis of artificial speech. He considered the injection transducer to be directly placed on the throat or in front of the mouth which was equipped with separate noise and pulse generation mechanisms to produce voiced and unvoiced phonemes.

Based on the classification in the preamble of this section, MacLeod's proposed system was a substitutive approach which converted a speaker's silently mouthed words into synthesized audible speech. Other later authors mainly considered supplementary ultrasonic excitation, mostly for speech recognition. (Tosaya & Sliwa, 2002; 1999) patented a system which applied ultrasonic signal injection to the vocal tract to make the task of audible voice recognition more robust. Their system was proposed to enhance or replace the natural excitation with an artificial excitation for which ultrasound was considered an option. The injection points for the artificial excitation were proposed to include: outside and within the mouth, nasal passage and on the neck.

Another instance of ultrasonic speech implementation was proposed by (Lahr, 2002). He considered the ultrasonic output of the VT as the third mode of a trimodal voice recognition system whose other two modes were audible voice and images of the lips, tongue and the teeth. In addition to greater transcription accuracy in the recognition task, the system was claimed to be capable of audible speech production when the speaker did not use vocal fold vibration and just shaped the VT in positions associated to several different voices. He elected to use the neck and mouth as possible injection points of 28 to 100 kHz excitations. He also stated that wearing a neck device was usually uncomfortable so he focused on signal injection over the lips where the mouth and teeth opening permitted the signal to penetrate in the VT. The ultrasonic output of his system was finally demodulated to the audible range and used directly as an input channel to a recognition system.

Another implementation was reported by (Douglass, 2006), who used ultrasonic excitation to add value in improving the reliability of speech recognition. His excitation points were below the chin, on the neck, in front, and inside of the mouth. He proposed employing the same means of demodulating commonly used in radio broadcasting for the output ultrasonic signal.

4.3 Necessary considerations for implementation

There are several considerations which are necessary for implementation of an ultrasonic speech system. These considerations include, signal injection points, excitation waveforms, feature extraction method and hardware setup.

As stated in section 4.2, in spite of its various applications, ultrasonic speech has been a little researched area and there have been few cases of attempts of implementation. One of the reasons for unpopularity might be problems associated with signal injection to the vocal tract. The choice of injection position has a great impact on system design. Ultrasound, as we have observed in section 3.1, reflects back almost totally from the air-tissue interface. Another strong reflecting boundary is the bone/soft tissue interface. The bone is normally avoided in ultrasound propagation, because it distorts the ultrasonic beam (Zangzebki, 1996) (so we will not consider placing the transducer on the jaw or skull bones in this chapter). Consequently, injecting the signal through the bone or when the signal is going to face an air-tissue interface before entering the VT are not promising options.

Nevertheless, the task of signal injection is possible via some considerations to prevent or compensate for injection problems. Possible injection points introduced by previous implementations include the throat, on the neck, against the cheek, in the nasal cavity, inside and in front of the mouth. Each of these injection points imposes special considerations to fulfil the task of augmentation of an ultrasonic excitation to the VT.

As an example, for signal injection over the neck skin which has been used by (Lahr, 2002; MacLeod, 1987; Tosaya & Sliwa, 2002), the ultrasound wave propagates from the transducer to the air gap between the transducer and skin. As we have previously observed, this air/tissue boundary totally reflects the signal back. We can compensate for the effect of the reflection by using a coupling gel on the skin to eliminate the air from the transducer/skin interface. The signal entering the skin passes the tissue and encounters another tissue/air boundary before being able to enter the vocal tract where it will almost totally reflect back. So to consider signal injection over the neck skin we may need to apply the injection where the tissues are relatively thin to minimize reflection effects over the thin boundary. Another convenient option is signal injection in front of the mouth.

Excitation signal waveform design is another task which could simplify and optimize the operation of the system. Another brain-storming task is the down conversion of ultrasonic output and extraction of features which will be used for the reconstruction or recognition of audible speech. Although some of the previously mentioned implementations have considered the demodulation of ultrasonic speech to gain the audible equivalent, when the resulting converted signal is going to provide features to produce audible speech, the design of ultrasonic speech systems will require greater attention. This chapter addresses a solution to this issue by mathematically proving the possibility of linear predictive analysis (LPA) of ultrasonic speech. LPA is one of the strong feature extraction facilities based on a linear source-filter model of speech production. Extension of LPA to the ultrasonic domain will significantly simplify processing and analysis requirements in the audible domain.

The choice of hardware components in any ultrasonic system is another implementation consideration. Transducers are the core of a typical ultrasonic set up, fulfilling the task of transmit and receive, but ultrasonic system set up comprises several other hardware components including a signal generator to supply input energy to the transmitting transducer, and a data acquisition system to capture the signals for analysis.

5. Human speech production anatomy and physiology

The human speech production apparatus is well designed for the task of generating, modulating, and projecting intelligible sound. Controlled, in part by the Broca nucleus in the frontal cortex and Wernicke nucleus in the temporal cortex of the brain, the muscles controlling lung exhalation, glottal tension, epiglottis, tongue, throat and lip position, must work in concert to create and modulate the sounds that make up language.

Although speech can be considered as simply as a set of complex waveforms, and indeed sinewave speech can be created from simple waveforms (McLoughlin, 2009), it is in reality a complex and rich set of auditory symbols differentiated through several production mechanisms. These are illustrated in figure 3, and include the following:

- airflow from the lungs, either restricted, diverted through the nasal passages, around the tongue, through the lips or teeth, modulated in speed and intensity, or blocked momentarily, as in a plosive sound like /p/. It is the job of the lungs to provide the airflow, and to modulate its intensity (although the glottis and lips can both be used to block airflow for a time).
- pitch comes from the vibration of the flap-like vocal cords in the glottis, induced by airflow from the lungs. As some muscles in the glottis tauten, the glottal opening narrows and the vibration consequently increases in frequency. Pitch not only provides the characteristic frequency of our speech, but contributes a lexical meaning in several languages, particularly Chinese. Perhaps the most important role of pitch, which is similar in many ways to a periodic pulse train, is to resonate through the vocal tract.
- vocal tract geometry dictates the resonance patterns produced by the excitation. A pitch train flowing through the VT causes these resonances which affect the frequency of the sound exiting the tract in much the same way as most wind instruments operate.

Consider further this analogy with a wind instrument: a trumpet player relies upon a mouthpiece which, when blown, acts with the lips to produce a buzzing sound. This takes the place of the glottis in the speech production mechanism (and both examples require lungs to make the air move in the first place). The annoying buzzing sound from a trumpet mouthpiece, when fed through the smooth tubes of a trumpet, results in a beautiful resonant horn sound. Pressing or releasing the trumpet valves (keys) selects the tubes that the air passes through, resulting in different notes being played. Similarly, the glottal vibration is modified by the vocal tract to produce speech sounds. Changing the geometry of the vocal tract under muscular control changes the sounds produced in speech (McLoughlin, 2009).

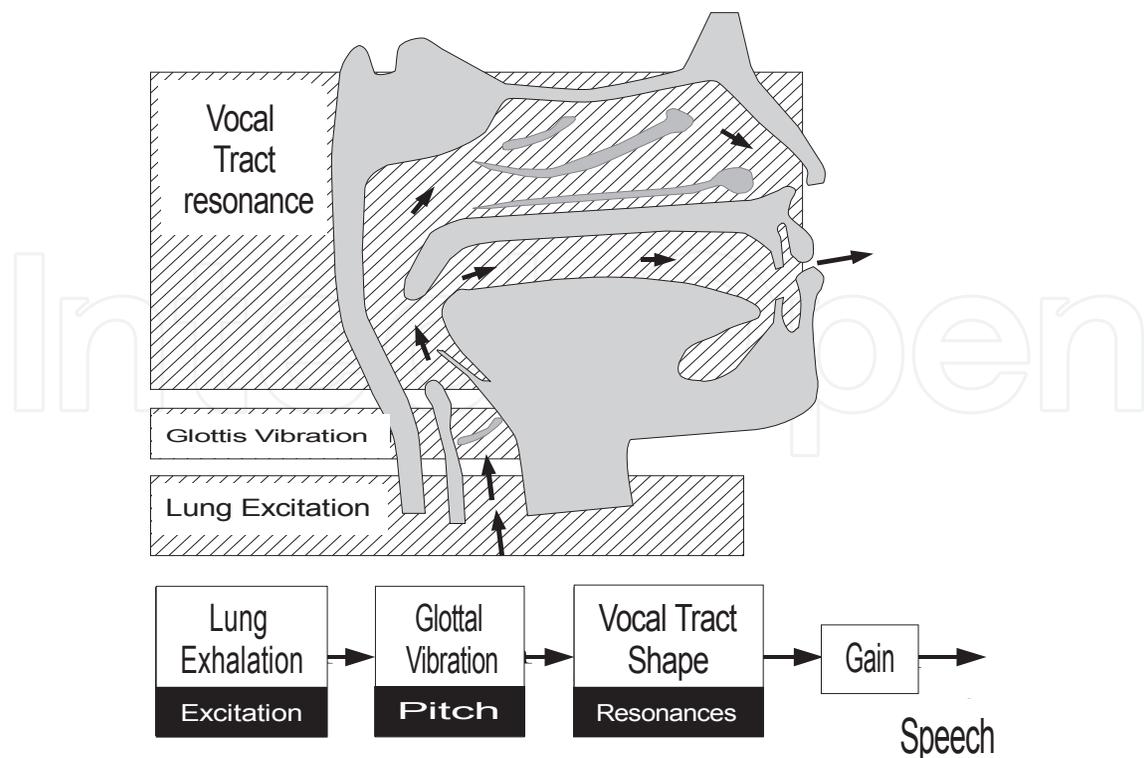


Fig. 3. A cut-away diagram of the human speech production mechanism, namely the human head (top), along with a block diagram representation below, showing lung excitation causing pitch to be produced by the glottis, acted upon by the vocal tract, and emitted from the mouth and nose

In speech, pitch is not present in all sounds: the vowel /a/ is voiced, meaning that it contains pitch, whereas the letter /f/ is unvoiced – meaning there is no pitch, so the sound is all lung excitation plus vocal tract shape. However all vowels are voiced, as are many consonants.

In ultrasonic speech production, an ultrasonic pulse-train usually replaces the pitch component generated by the glottis. All other articulators remain: the lungs still exhale, and provide airflow for the quiet unvoiced sounds (which are around 16dB quieter than voiced sounds). The tongue, lips and throat muscles still act together and the human brain can still direct the voice production apparatus to form words, as if whispering (which is naturally unvoiced). The main difference being that the pulse-producing glottis does not resonate.

Finally, understanding the speech production mechanism led many researchers to adopt a source-filter model for speech. This model separates the sound source (lung and glottis), from the filter (vocal tract), and assumes that these two parts are independent, but when directed by the brain to act in concert, produce the required sounds. Almost all modern speech analysis and processing systems rely heavily upon the source-filter model, and in particular assume that the filter part of the model can be represented by a linear polynomial function. It is this important relationship that we aim to establish for the case of LF ultrasonic speech.

6. Modelling ultrasonic speech process

Linear partial differential equations (PDEs) are the basic descriptors of linear systems, as a consequence of allowance to the principle of superposition (Coleman, 2005). Well known for benign impulse response and convolutional characteristics, linear time invariant (LTI) systems theory has underpinned the source-filter model of speech production for decades. The aim of this section is to derive a linear model for the propagation of ultrasonic signals through the vocal tract. We have seen in the previous sections that the assumptions of lossless propagation and ideal gas behaviour are plausible for small amplitude LF ultrasound propagation within the vocal tract. We commence our modelling from basic acoustic equations and apply these.

6.1 Mathematical description of ultrasonic propagation in the VT

The theory of acoustics stems from four main PDEs based on the conservation of mass, momentum and energy and also equations of the state of the medium (Blackstock, 2000), valid in three dimensional space over the frequency range of sound waves (including infrasound, audio and ultrasound). These equations are generally not linear but they are linearized in acoustics under several simplifying assumptions (Reynolds, 1981) and lead to the facilities of the theory of linear acoustics. We have theoretically described these assumptions earlier but will now review them mathematically before building on them further.

The first assumption is to consider ultrasonic wave propagation to be an adiabatic (lossless) phenomenon. We observed that the main causes of attenuation in ultrasound frequencies in a fluid medium are heat conduction, relaxation and viscosity. We then observed in section 3.2.1 that the effect of this attenuation in the frequency range of our application is negligible. So the process could be considered lossless (adiabatic) in which case, the equation of energy conservation will not be necessary (Blackstock, 2000).

The remaining equations are conservation of momentum (9) and mass (10) and equations of state of the gas. These equations describe the evolution of pressure \tilde{p} and particle velocity vector $\tilde{\mathbf{u}}$ as functions of time t and three dimensional coordinates, $\mathbf{r} = [x \ y \ z]$. The general form of these equations is as stated below (Reynolds, 1981) where $\tilde{\rho}$ is the density, μ and λ are viscosity coefficients of the medium and F is the external excitation force.

$$\tilde{\rho} \left(\frac{\partial \tilde{\mathbf{u}}}{\partial t} + \nabla \cdot (\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}) \right) = -\nabla \tilde{p} + (2\mu + \lambda) \nabla (\nabla \cdot \tilde{\mathbf{u}}) - \mu \nabla \times \nabla \times \tilde{\mathbf{u}} + F \quad (9)$$

$$\frac{\partial \tilde{\rho}}{\partial t} + \nabla \cdot (\tilde{\rho} \tilde{\mathbf{u}}) = 0 \quad (10)$$

Equation (9) includes the divergence of a dyadic product which is defined as:

$$\mathbf{v} \otimes \mathbf{u} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} = \begin{bmatrix} v_1 u_1 & v_1 u_2 & v_1 u_3 \\ v_2 u_1 & v_2 u_2 & v_2 u_3 \\ v_3 u_1 & v_3 u_2 & v_3 u_3 \end{bmatrix} \quad (11)$$

where u_i is the i^{th} element of the vector \mathbf{u} .

The system (9-10) is completed by the equation of state that gives the pressure as a function of the density and temperature. When the flow is adiabatic in a gas, that is, no heat is transferred to or from the gas, and is reversible, that is, the flow conditions can return to

their original values, the pressure is a function of the density only (Fahy, 2001), and the equation of state of the gas reduces to:

$$\tilde{p} = \tilde{p}(\tilde{\rho}) \quad (12)$$

Considering the equation of conservation of momentum (9), with adiabatic and reversible wave deformation in the medium, the next assumption is irrotational flow, $\nabla \times \tilde{\mathbf{u}} = 0$. This assumption has been somehow challenged by the existence of rotational flows in turbulent and jet flows in the classical linear modelling of audible sound propagation in the vocal tract during articulation of unvoiced utterances.

Due to the work of (Lighthill, 1952) and (Goldstein, 1984) the production of turbulent flow is governed by nonlinear equations of acoustics but once fully developed, we can describe its propagation as irrotational, governed by equations of linear acoustics (Crocker, 2007). We have conventionally used this assumption for the audible case, transferring the non-linearity of turbulent flow production to the source and dealing with the VT as a linear filter in the conventional source-filter modelling of the speech production system (Sinder, 1999). The same considerations apply to the ultrasonic range and make the assumption of $\nabla \times \tilde{\mathbf{u}} = 0$ a plausible statement.

The next step is to consider the effects of viscosity. Based on the discussions of section 3.2.2 about negligible dispersive effects of viscosity for frequencies below 50 MHz and referring to section 3.2.1 about values of visco-thermal absorption coefficient of the air in the frequency range of the current application, we can consider μ and λ to be very small, to neglect the effects of viscosity for LF ultrasound propagating in the air. We may now rewrite (9) in a clearer notation of (13) for each j from 1 to 3 as:

$$\tilde{\rho} \left(\frac{\partial \tilde{u}_j}{\partial t} + \sum_{i=1}^3 \frac{\partial (\tilde{u}_i \tilde{u}_j)}{\partial x_i} \right) + \frac{\partial \tilde{p}}{\partial x_j} = F_i \quad (13)$$

Considering Small disturbances in pressure and density we will have (14, 15) where $p_0, \rho_0, \mathbf{u}_0$ are attributes of the medium at equilibrium state which are actually the time averages of $\tilde{p}, \tilde{\rho}$ and $\tilde{\mathbf{u}}$ respectively. "Acoustic pressure" p is introduced here then as the small variations of pressure around the equilibrium value p_0 .

$$\tilde{p} = p_0 + p; \quad \tilde{\rho} = \rho_0 + \rho; \quad \tilde{\mathbf{u}} = \mathbf{u}_0 + \mathbf{u} \quad (14)$$

$$\frac{\partial p_0}{\partial t} = 0; \quad \frac{\partial \rho_0}{\partial t} = 0; \quad \frac{\partial \mathbf{u}_0}{\partial t} = 0 \quad (15)$$

Assuming the homogeneous (16) medium initially at rest (17):

$$\nabla p_0 = 0 \quad ; \quad \nabla \rho_0 = 0 \quad (16)$$

$$\mathbf{u}_0 = 0 \quad (17)$$

And manipulating conditions of (14-17) in (13), the linear equation of conservation of acoustic momentum for a lossless homogeneous medium initially at rest is derived for ultrasonic propagation inside the vocal tract by (18):

$$\rho_0 \frac{\partial \mathbf{u}}{\partial t} + \nabla p = F \quad (18)$$

For the equation of conservation of mass (10), using the above assumptions of homogeneous medium, small disturbances and medium at rest (14-17), we can determine the following:

$$\frac{\partial \rho}{\partial t} + \rho_0 \nabla \cdot \mathbf{u} = 0 \quad (19)$$

The equation of state for an ideal gas states that:

$$\frac{p}{\rho} = \frac{\partial \tilde{p}}{\partial \tilde{\rho}} = c^2 \quad (20)$$

Where c is the speed of sound. The dispersive effects of air medium are discarded in (20) based on the discussions of section 3.2.2. Taking the derivative of (20) with respect to time, we will have:

$$\frac{\partial p}{\partial t} = c^2 \frac{\partial \rho}{\partial t} \quad (21)$$

Substituting (21) in (19) we would reach to the conservation of mass equation for ultrasonic propagation in the vocal tract:

$$\frac{1}{c^2} \frac{\partial p}{\partial t} + \rho_0 \nabla \cdot \mathbf{u} = 0 \quad (22)$$

We would rewrite (18,22), i.e. lossless linear acoustic equations in (23,24) as the basic equations of ultrasound propagation in the vocal tract where p is the acoustic pressure and \mathbf{u} is the acoustic velocity vector, ρ_0 is the static mass density of the medium and K is the adiabatic bulk modulus of the air:

$$\rho_0 \frac{\partial \mathbf{u}}{\partial t} + \nabla p = F \quad (23)$$

$$\frac{\partial p}{\partial t} + K \nabla \cdot \mathbf{u} = 0 \quad (24)$$

As observed mathematically, the derivation of ultrasonic wave propagation in the vocal tract, with the simplifying assumptions which we have described in detail, has led to equations (23), (24) which are the general equations of linear acoustics, now applicable for ultrasonic propagation through the vocal tract.

6.2 Vocal tract transfer function for ultrasonic speech

In our approach to derive a linear model, in this section the VT transfer function is determined using the functional transformation method (FTM) which converts the linear PDEs to algebraic equations including boundary and initial conditions, similarly to Laplace transformation in ordinary PDEs (Rabenstein, 1999).

Combining (23) and (24) yields the wave equation for $p(t, \mathbf{r})$ and $\mathbf{u}(t, \mathbf{r})$:

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} - \nabla^2 p = 0 \quad ; \quad \frac{1}{c^2} \frac{\partial^2 \mathbf{u}}{\partial t^2} - \nabla^2 \mathbf{u} = 0 \quad (25)$$

where $\mathbf{r} = [x \ y \ z]$ is the three dimensional coordinates vector and c is the speed of sound.

For audible sound production, since the cross section of the VT is small compared to the wavelength, the wave can propagate along the tract axis and we can model the VT simply as a single narrow tube. However the smaller wavelength of ultrasound means the wave can propagate across the width of the tract and the resulting cross modes require (25) solving in three dimensions. Thus the task of derivation of the three dimensional VT transfer function may not be as simple as the one dimensional wave equation for audible sound. We are considering the placement of the source in front of the mouth, however the general method is applicable to other injection positions.

Representing VT volume as Ω and its boundary as Γ being comprised of boundaries Γ_1 (the glottis), Γ_2 (VT walls) and Γ_3 (the mouth), having $f(r, t)$ to be the ultrasonic excitation source placed in front of the mouth, the general boundary and initial conditions of ultrasonic wave propagation in the VT can be found, with $Z(\mathbf{r})$ being the impedance of the VT and closed glottis walls, as:

$$\left\{ \begin{array}{ll} p(0, \mathbf{r}) = 0; & \mathbf{r} \in \Omega; \quad \text{medium initially at rest} \\ \frac{\partial p}{\partial t}(0, \mathbf{r}) = p_0(\mathbf{r}); & \mathbf{r} \in \Omega \\ (n \cdot (\frac{1}{\rho_0} \nabla) + \frac{1}{Z(\mathbf{r})} \frac{\partial}{\partial t}) p(\mathbf{r}, t) = 0 & \mathbf{r} \in \Gamma_1, \Gamma_2 \\ n \cdot \nabla p(\mathbf{r}, t) = f(\mathbf{r}, t) & \mathbf{r} \in \Gamma_3 \end{array} \right. \quad (26)$$

Defining linear differential operators: $D_t = \frac{\partial^2}{\partial t^2}$, $D_r = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, we can rewrite (25) for pressure as:

$$\frac{1}{c^2} D_t(p(\mathbf{r}, t)) - D_r(p(\mathbf{r}, t)) = 0 \quad (27)$$

Taking the Laplace transform of (27) and considering the initial conditions of (26), we convert differential operator $D_t(p)$ to the algebraic form:

$$\left\{ \begin{array}{l} \frac{1}{c^2} s^2 P(\mathbf{r}, s) - \frac{1}{c^2} p_0(\mathbf{r}) - D_r(P(\mathbf{r}, s)) = 0 \end{array} \right. \quad (28.a)$$

$$\left\{ \begin{array}{l} (n \cdot (\frac{1}{\rho_0} \nabla) + \frac{1}{Z(\mathbf{r})} s) P(\mathbf{r}, s) = 0 \end{array} \right. \quad \mathbf{r} \in \Gamma_1, \Gamma_2 \quad (28.b)$$

$$\left\{ \begin{array}{l} n \cdot \nabla P(\mathbf{r}, s) = F(\mathbf{r}, s) \end{array} \right. \quad \mathbf{r} \in \Gamma_3 \quad (28.c)$$

$P(\mathbf{r}, s)$ is the Laplace transform of $p(\mathbf{r}, t)$. Next we seek another transform T which can convert the spatial differential operator D_r to algebraic equations. Lacking a general transform similar to the Laplace transform in the spatial domain, the spatial Sturm-Liouville transform (SLT) (Rabenstein, 1999) is applied:

$$T(P(\mathbf{r})) = P_T(\beta_k) = \iiint_{\Omega} P(\mathbf{r}) \cdot K(\mathbf{r}, \beta_k) dV \quad (29)$$

The dependence upon Laplace transform parameter (s) is omitted for convenience from this point on (so $P(\mathbf{r}, s)$ is written as $P(\mathbf{r})$ for instance). The aim is to evaluate the kernel function $K(\mathbf{r}, \beta_k)$ so that:

$$T(D_r\{P(\mathbf{r})\}) = \beta_k^2 T(P(\mathbf{r})) + \Phi_b(\Gamma) \quad (30)$$

Where β_k is a scalar coefficient and $\Phi_b(\Omega)$ is a function which depends on the boundary conditions of the problem. To reach this goal, we first multiply (28.a) by $K(\mathbf{r}, \beta_k)$.

$$K(\mathbf{r}, \beta_k) \left\{ \frac{1}{c^2} s^2 P(\mathbf{r}) - \frac{1}{c^2} p_0(\mathbf{r}) - D_r \{P(\mathbf{r})\} = 0 \right\} \quad (31)$$

Next we take the integral $\iiint_{\Omega} dV$, dV is the volume element.

$$\frac{1}{c^2} s^2 \iiint_{\Omega} P(\mathbf{r}) K(\mathbf{r}, \beta_k) dV - \frac{1}{c^2} \iiint_{\Omega} p_0(\mathbf{r}) K(\mathbf{r}, \beta_k) dV - \iiint_{\Omega} K(\mathbf{r}, \beta_k) D_r \{P(\mathbf{r})\} dV = 0 \quad (32)$$

Referring to the definition of the SL transform (29), (32) yields:

$$\frac{1}{c^2} s^2 P_T(\beta_k) - \frac{1}{c^2} p_{0T}(\beta_k) - \iiint_{\Omega} K(\mathbf{r}, \beta_k) D_r \{P(\mathbf{r})\} dV = 0 \quad (33)$$

Considering $D_r = \nabla^2$ and by Green's theorem (Rabenstein, 1999), the integral in (33) is:

$$\begin{aligned} T\{D_r \{P(\mathbf{r})\}\} &= \iiint_{\Omega} K(\mathbf{r}, \beta_k) D_r \{P(\mathbf{r})\} dV = \iiint_{\Omega} K(\mathbf{r}, \beta_k) \nabla^2 P(\mathbf{r}) dV \\ &= \iiint_{\Omega} [P(\mathbf{r}) \nabla^2 K(\mathbf{r}, \beta_k)] dV \\ &+ \iint_{\Gamma} K(\mathbf{r}, \beta_k) \nabla P(\mathbf{r}) \cdot n d\sigma - \iint_{\Gamma} P(\mathbf{r}) \nabla K(\mathbf{r}, \beta_k) \cdot n d\sigma \end{aligned} \quad (34)$$

$n d\sigma$ is the surface element. Comparing (34), and (30), the first integral in the right hand side of (34) should be converted to a multiplicand of $T\{P(\mathbf{r})\}$ (29). The second integral uses the values of $\nabla P(\mathbf{r})$ on the boundary Γ , which we have by the boundary conditions of (26). The last term is unwanted because we do not have the value of $\{P(\mathbf{r})\}$ over the boundary so we define kernel $K(\mathbf{r}, \beta_k)$ to fulfil the following requirements as:

$$\begin{cases} \nabla^2 K(\mathbf{r}, \beta_k) = \beta_k^2 K(\mathbf{r}, \beta_k) \\ \nabla K(\mathbf{r}, \beta_k) = 0 \quad \mathbf{r} \in \Gamma \end{cases} \quad (35)$$

Equation (35) is the well known Helmholtz equation (Blackstock, 2000) and its general solution relies strongly to the geometry Ω . Values of $K(\mathbf{r}, \beta_k)$, β_k are Eigen functions and Eigen values of the operator $D_r = \nabla^2$ (Rabenstein, 1999). We then substitute the results in (33):

$$\frac{1}{c^2} s^2 P_T(\beta_k) - \frac{1}{c^2} p_{0T}(\beta_k) - \iiint_{\Omega} \beta_k^2 K(\mathbf{r}, \beta_k) P(\mathbf{r}) dV = \iint_{\Gamma} \{K(\mathbf{r}, \beta_k)\} \nabla P(\mathbf{r}) \cdot n d\sigma \quad (36)$$

Referring to the definition of SLT (29) and substituting the values of $\nabla P(\mathbf{r})$ from boundary conditions (28.a,b), we may rewrite (36) as:

$$\begin{aligned} \left(\frac{1}{c^2} s^2 - \beta_k^2\right) P_T(\beta_k) &= \frac{1}{c^2} p_{0T}(\beta_k) - s \rho_0 \iint_{\Gamma_{1,2}} \frac{P(\mathbf{r})}{Z(\mathbf{r})} K(\mathbf{r}, \beta_k) d\sigma + \iint_{\Gamma_3} K(\mathbf{r}, \beta_k) F(\mathbf{r}) d\sigma = \\ &\frac{1}{c^2} p_{0T}(\beta_k) - s \rho_0 G\left(\frac{P(\mathbf{r})}{Z(\mathbf{r})}, \Gamma_{1,2}, \beta_k\right) + G(F(\mathbf{r}), \Gamma_3, \beta_k) \end{aligned} \quad (37)$$

Equation (37), where $G(F(\mathbf{r}), \Gamma_j, \beta_k) \triangleq \iint_{\Gamma_j} K(\mathbf{r}, \beta_k) F(\mathbf{r}) d\sigma$, is the general equation relating the output $P_T(\beta_k)$ of the VT to the input $F(\mathbf{r})$ and initial and boundary conditions.

Considering hard walls for both the vocal tract and closed glottis, $Z(\mathbf{r}) \rightarrow \infty$ (based on the impedance values of the soft tissue in section 3.1) and $p_{0T}(\beta_k) = 0$, i.e. zero initial conditions,

and $F(r, s) = F(s)$ meaning that the ultrasound source has uniform spatial distribution pattern, which is a plausible simplification we have:

$$G(F(\mathbf{r}), \Gamma_3, \beta_k) = F \iint_{\Gamma_3} K(\mathbf{r}, \beta_k) d\sigma \quad (38)$$

And consequently:

$$P_T(\beta_k) = \frac{(c^2 F) \cdot \iint_{\Gamma_3} K(\mathbf{r}, \beta_k) d\sigma}{s^2 - c^2 \beta_k^2} \quad (39)$$

Since $P_T(\beta_k) = T\{P(\mathbf{r})\}$, we need to take the inverse SL transform (Rabenstein, 1999) to reach $P(\mathbf{r})$.

$$T^{-1}(P_T(\beta_k)) = P(\mathbf{r}) = \sum_{k=1}^{\infty} \frac{1}{N_k} P_T(\beta_k) K(\mathbf{r}, \beta_k) \quad (40)$$

$$N_k = \iiint_{\Omega} K^2(\mathbf{r}, \beta_k) dV$$

$P(\mathbf{r}) = P(\mathbf{r}, s)$ is the Laplace transform of $p(\mathbf{r}, t)$. Using simplifications of (39), (40) becomes:

$$P(\mathbf{r}) = F c^2 \left[\sum_{k=1}^{\infty} \frac{1}{N_k} \left\{ \iint_{\Gamma_3} K(\mathbf{r}, \beta_k) d\sigma \right\} K(\mathbf{r}, \beta_k) \right] \quad (41)$$

And consequently we will reach the transfer function of vocal tract for ultrasonic speech:

$$H(\mathbf{r}) = \frac{P(\mathbf{r})}{F} = c^2 \left[\sum_{k=1}^{\infty} \frac{1}{N_k} \left\{ \iint_{\Gamma_3} K(\mathbf{r}, \beta_k) d\sigma \right\} K(\mathbf{r}, \beta_k) \right] \quad (42)$$

$H(\mathbf{r})$ is the three dimensional transfer function of the vocal tract when excited in front of the mouth which explicitly is a function of \mathbf{r} but in its formation, the integrals were on the geometry of volume Ω and its boundaries Γ , so $H(\mathbf{r})$ is strongly relied on the definition of the geometry. Thus the three-dimensional wave equation applied to the near-audio ultrasonic speech, with several realistic assumptions as described, yields the linear transfer function (42).

6.3 Linear source filter model for ultrasonic speech

Showing the Laplace transform parameter (s) again - which we had omitted in our equations up to the point for simplicity - we recall that $H(\mathbf{r})$ was actually $H(\mathbf{r}, s)$, the Laplace transform of $h(\mathbf{r}, t)$. If sampling time intervals are small enough to consider the VT shape pseudo-static, a system with transfer function $h(\mathbf{r}, t)$ will be an LTI system, leading to a convolutional relation between its output and input as (43). So $h(\mathbf{r}, t)$ can be considered as a linear time-invariant (LTI) filter for small time intervals and by the benefit of LTI systems, the conventional source-filter model of audible speech can be extended to cover ultrasonic speech production.

$$p(\mathbf{r}, t) = h(\mathbf{r}, t) * f(t) \quad (43)$$

The classical source-filter modelling of VT enjoys independence between source and filter. In the case of ultrasonic speech, the source and the filter are intrinsically independent.

7. Extension of LPA to the analysis of ultrasonic speech

In the previous section, linear source filter model of speech production was mathematically proven to be valid for ultrasonic speech. Linear source filter modelling of ultrasonic speech is the basis of linear predictive analysis as a powerful feature extraction method as will be observed in this section.

The Z transform of $h(\mathbf{r}, t)$, can be described as an IIR filter as in (44).

$$H(\mathbf{r}, z) = \frac{\sum_{i=1}^N b_i(\mathbf{r})z^{-i}}{1 + \sum_{i=1}^M a_i(\mathbf{r})z^{-i}} \quad (44)$$

There is a need to inspect the dependence of $H(\mathbf{r}, z)$ on coordinates vector \mathbf{r} more carefully. The VT is a resonant cavity and at ultrasonic frequencies will have cross modes of resonance. If the excitation signal is a sine function of the same frequency of one of the modes of the resonance, a standing wave of that frequency will form and as a consequence of linearity, the output wave at any point, except nodes, will have the same frequency as the input. The impulse function is the integral sum of an infinite number of sine waves in the time domain. As another consequence of LTI systems, the response of the VT to the impulse will be the summation of its output to sine waves of all frequencies including all its resonances with different amplitudes. Accordingly although the transfer function would have different values in different \mathbf{r} , it will have the same set of common poles as the resonances of the tract. These common resonances can be calculated with several methods as per (Haneda & Makino et al., 1994).

Linear predictive analysis utilizes the autoregressive (all pole) representation of the transfer function of VT and provides the procedures to evaluate the coefficients of the denominator. The same procedure can be applied to the Z transform of the VT transfer function in (44) which as the transfer function of a minimum phase system, has both poles and zeros inside the unit circle and can be represented as an all pole transfer function, with any zeros being approximated by additional poles (Rabiner & Schafer, 1978).

8. Open research questions

This chapter has presented a mathematical model for ultrasound propagation in the vocal tract and has proven the possibility of application of linear predictive analysis to the ultrasonic speech. The source-filter model of speech production and LPA are the basic building blocks of audible speech processing. Expanding their implementation to ultrasonic speech is the major basis of implementation of this technology. Having the findings of this chapter in hand, ultrasonic speech can begin to enjoy further research effort to reach a state of maturity.

For ultrasonic speech, an ultrasound excitation is injected into the vocal tract. The choice of optimum excitation point and excitation signal wave-form is a topic for further research. Based on the achievements of this chapter, the ultrasonic speech at the output of the mouth can be treated as the output of a LTI source-filter model and can be subjected to LPA analysis to retrieve a set of common poles of the transfer function. The extracted features, converted to a set of parameters, are suitable for production of audible speech. Efficient and accurate down-conversion is also a topic of further research which involves the choice of suitable deterministic or statistic conversion methods.

Finally, as ultrasonic speech involves long term exposure to ultrasound frequencies below 100 kHz, medical standards in place relating to the health effects of the technology need to be assessed and possibly revised as a pre-condition to widespread adoption.

9. Conclusion

This chapter has presented ultrasonic speech as a novel application of ultrasound in speech augmentation. Ultrasonic speech, operating by replacing the natural excitation in audible speech with an LF ultrasonic signal, has applications in speech augmentation for the speech rehabilitation and secure communications communities. This chapter has studied the requirements in modelling ultrasonic speech as a linear system of sound propagation and has proven that LPA, a major tool in the analysis of normal speech, is also extendible to ultrasonic speech.

In pursuing this aim, we first introduced the attributes of ultrasonic propagation in a linear lossless gas medium. We observed that if the sound propagation is an adiabatic procedure and the gas obeys the ideal gas law and with small disturbances in the medium as a result of wave propagation, the gas medium can be considered a linear lossless medium for ultrasound propagation. We then discussed deviations of these conditions for ultrasound propagation in the air medium.

Subsequently, LF ultrasound was introduced, and the impacts of the deviations of linear acoustic behaviour were numerically analyzed for propagation of low frequency ultrasound in the vocal tract. Then we considered the application of LF ultrasound in speech augmentation and discussed the aspects of system design which seek more attention. By a review of previous implementations, we investigated how they had addressed these aspects including the injection points and methods of down-conversion to audible domain.

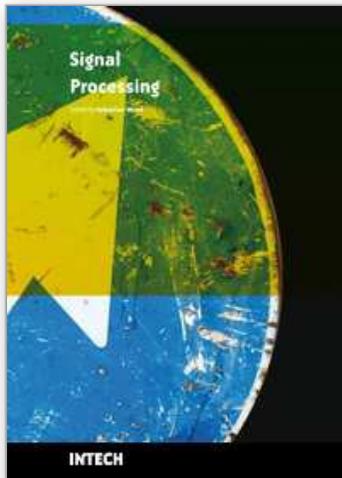
Afterwards we considered the physiology and anatomy of the human speech production mechanism and how we can substitute the natural excitation with an ultrasonic waveform in speech augmentation. We also stated that the ultrasonic excitation could be applied as a supplement to natural excitation to provide additional data for speech processing applications. The chapter then demonstrated a linear modelling scheme in addition to the fact that speech LPA tools can be extended to sound propagation at lower ultrasonic frequencies. Starting with basic wave equations, and making several simplifying assumptions such as rigid walls for closed glottis and VT, relatively small signal disturbance, and a spatially flat (uniform) excitation source, the VT has been shown to be LTI with the transfer function in the form of a pole-zero IIR filter. By means of this derivation, the conventional source-filter model was proven to be extendable for an ultrasonic speech production system, and thus the powerful tools of LPA can be used.

In this chapter we have tried to bridge from audible speech processing methods to ultrasonics by mathematically and physically demonstrating that the extension of principles of audible speech processing to the analysis of ultrasonic speech is plausible. This significantly simplifies ultrasonic speech processing. The currently neglected area of LF ultrasonics research in speech analysis and processing can now be explored with relative ease. Further research effort is necessary, and welcomed in this area, as it moves toward further maturity and future real-life applications.

10. References

- Akerman, M. A.; C. W. Ayers & H. D. Haynes (1994). Ultrasonic speech translator and communications system, United States Patent and Trademark Office, No. 5539705, 1996, United States.
- Avallone, E. A.; T. Baumeister; A. Sadegh & L. S. Marks (2006). *Marks Standard handbook for mechanical engineers*, McGraw-Hill Professional.
- Bauer, H. J. (1965). Theory of relaxation phenomena in gases, *Physical acoustics*, Vol. IIA.
- Begault, D. R. (1994). *3-D sound for virtual reality and multimedia*, Academic Press.
- Blackstock, D. T. (2000). *Fundamentals of physical acoustics*, Wiley Interscience.
- Blitz, J. (1967). *Fundamentals of ultrasonics*, Butterworth and Co.
- Bühler, O. (2006). *A Brief Introduction to classical, statistical, and quantum mechanics*, American Mathematical Society.
- Caruthers, J. W. (1977). *fundamentals of marine Acoustics*, Elsevier.
- Clark, C. W. (2004). Baleen whale infrasonic sounds: Natural variability and function, *Journal of Acoustical Society of America*, Vol. 115, No. 5, pp. 2554-2554.
- Coleman, M. P. (2005). *An introduction to partial differential equations with MATLAB*, CRC Press.
- Crocker, M. J. (1998). *Handbook of acoustics*, Wiley Interscience.
- Crocker, M. J. (2007). *Handbook of noise and vibration control*, John Wiley and Sons.
- David, J. & N. Cheeke (2002). *Fundamentals and applications of ultrasonic waves*, CRC press LLC.
- Dean, E. A. (1979). Atmospheric effects on the speed of sound. Technical report of Defense Technical Information Center.
- Douglass, B. G. (2006). Apparatus and method for detecting speech using acoustic signals outside the audible frequency range, United States Patent and Trademark Office, No. US 200710276658, United States.
- Ensminger, D. (1988). *Ultrasonics, fundamentals, technology, applications*, Marecel Dekker.
- Fahy, F. (2001). *Foundations of Engineering Acoustics*, Elsevier.
- Goldstein, M. (1984). *Aeroacoustics*, McGraw Hill.
- Haar, G. (1999). Therapeutic ultrasound, *European Journal of Ultrasound*, Vol. 9, No. 1, pp. 3-9.
- Halliday, D.; R. Resnick & J. Walker (2004). *Fundamentals of Physics*, John Wiley & Sons.
- Haneda, Y.; S. Makino & Y. Kaneda (1994). Common acoustical pole and zero modeling of room transfer functions. *IEEE trans. speech and audio proc.*, Vol. 2, No. 2.
- Harris, J. W.; W. Benenson; H. Stoecker & H. Lutz (2002). *Handbook of physics: with 797 illustrations*, Springer.
- Ikawa, M. (2000). *Partial differential equations*, American Mathematical Society.
- Ingard, U. (2008). *Notes on acoustics*, Infinity Science Press, LLC.
- Johnson, K. (2003). *Acoustic and auditory phonetics*, Blackwell Publishing.
- Karal, F. C. & J. B. Keller (1959). Elastic wave propagation in homogeneous and inhomogeneous media, *The journal of the acoustical society of America*, Vol. 31, No. 6, pp. 694-705.
- Karal, F. C. & J. B. Keller (1964). Geometrical theory of elastic surface-wave excitation and propagation, *The journal of the acoustical society of America*, Vol. 36, No. 1, pp. 32-40.
- Kelsey, C. A.; F. D. Minifie & T. J. Hixon (1969). Applications of ultrasound in speech research. *Journal of Speech and Hearing Research*, Vol. 12, pp. 564-575.

- Kyriakakis, C. (1998). Fundamental and technological limitations of immersive audio systems, *Proceedings of IEEE*, Vol. 86, No. 5.
- Lahr, R. J. (2002). Head-worn, trimodal device to increase transcription accuracy in a voice recognition system and to process unvocalized speech, United States Patent and Trademark Office, No. US 2002/0194005, 2002.
- Lempriere, B. M. (2002). *Ultrasound and elastic waves, Frequently asked questions*, Elsevier Science.
- Lenhardt, M. L.; R. Skellett; P. Wang & A. M. Clarke (1991). Human ultrasonic speech perception. *Science, New Series*, Vol. 253, No. 5015, pp. 82 - 85.
- Licker, M. (2002). *McGraw-Hill Dictionary of Scientific and Technical Terms*, McGraw-Hill Companies.
- Lighthill, M. J. (1952). On sound generated aerodynamically. i. general theory, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 211, No. 1107, pp. 564-587.
- MacLeod, N. (1987). Non-audible speech generation method and apparatus, U. S. Patent, No. 4821326, 1989.
- McLoughlin, I. V. (2009). *Applied speech and audio processing: with Matlab examples*, Cambridge University Press.
- Pozo, A. D. (2004). Laryngectomy speech enhancement using voice conversion techniques, Christ's College, University of Cambridge, M.Sc Thesis.
- Pujol, J. (2003). *Elastic wave propagation and generation in seismology*, Cambridge University Press.
- Rabenstein, R. (1999). Transfer function models for multidimensional systems with bounded spatial domains, *Mathematical and Computer Modelling of Dynamical Systems*, Vol. 5, pp. 259 – 278.
- Rabiner, L. R. & R. W. Schafer (1978). *Digital processing of speech signals*, Prentice-Hall.
- Rauch, J. (2008). *Hyperbolic partial differential equations and geometric optics*.
- Reynolds, D. D. (1981). *Engineering principles in acoustics*, Allyn and Bacon Inc.
- Sadd, M. H. (2005). *Elasticity: Theory, applications and numerics*, Academic Press.
- Serway, R. A. & J. W. Jewett (2006). *Principles of physics: a calculus-based text*, Thomson Brooks/Cole.
- Shawker, T. H. & B. C. Sonies (2005). Tongue movement during speech: a real-time ultrasound evaluation, *Journal of Clinical Ultrasound*, Vol. 12, No. 3, pp. 125 - 133.
- Simpson, J. A. & E. S. C. Weiner (1989). *The Oxford English Dictionary*, Clarendon Press.
- Sinder, D. J. (1999). Speech synthesis using an aeroacoustic fricative model, Graduate school- New Brunswick Rutgers, The State University of New Jersey, PhD thesis.
- Subramanian, C. V. (2006). *Practical ultrasonics*, Alpha Science.
- Szabo, T. L. (2004). *Diagnostic ultrasound imaging: inside out*, Academic Press.
- Tosaya, C. A. & J. W. Sliwa (2002, 1999). Signal Injection coupling into the human vocal tract for robust audible and inaudible voice recognition, United States Patent and Trademark Office, No. 7082395 & No. 6487531, United States.
- Watkinson, J. (1998). *The art of sound reproduction*, Focal Press.
- Webster, M. (1986). *Webster's ninth collegiate dictionary*, Springfield.
- Zangzebki, J. A. (1996). *Essentials of ultrasound physics*, Mosby, Elsevier.
- Zemlin, W. R. (1997). *Speech and hearing science anatomy and physiology*, Allyn and Bacon.



Signal Processing

Edited by Sebastian Miron

ISBN 978-953-7619-91-6

Hard cover, 528 pages

Publisher InTech

Published online 01, March, 2010

Published in print edition March, 2010

This book intends to provide highlights of the current research in signal processing area and to offer a snapshot of the recent advances in this field. This work is mainly destined to researchers in the signal processing related areas but it is also accessible to anyone with a scientific background desiring to have an up-to-date overview of this domain. The twenty-five chapters present methodological advances and recent applications of signal processing algorithms in various domains as telecommunications, array processing, biology, cryptography, image and speech processing. The methodologies illustrated in this book, such as sparse signal recovery, are hot topics in the signal processing community at this moment. The editor would like to thank all the authors for their excellent contributions in different areas of signal processing and hopes that this book will be of valuable help to the readers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Farzaneh Ahmadi and Ian McLoughlin (2010). The Use of Low-Frequency Ultrasonics in Speech Processing, Signal Processing, Sebastian Miron (Ed.), ISBN: 978-953-7619-91-6, InTech, Available from: <http://www.intechopen.com/books/signal-processing/the-use-of-low-frequency-ultrasonics-in-speech-processing>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen