EECS 225d: Midterm 1, March 6, 1995

Version with correct answers (not necessarily unique)

For multiple choice and true/false questions, also write a brief (1-2 sentence) explanation of why the answer is correct. The parenthesized numbers after each question number give the relative point value of each question.

1. (2) The lowest (frequency) mode of an acoustic tube open at both ends is

   a) higher than **X**
   b) equal to
   c) lower than

   the lowest node of an acoustic tube open at one and closed at the other end.

   A tube open at both ends will have zero pressure at those ends. For the lowest mode, this means that the tube length is half the wavelength. For the tube closed at one end and open at the other, the tube length is $\frac{1}{4}$ of the wavelength. This corresponds to a greater wavelength and a lower frequency. Thus a) is correct.

2. (2) The impedance of an acoustic tube

   a) increases
   b) stays the same
   c) decreases **X**

   as the mass density ρ decreases.

   Impedance Z is equal to ρc^2 and $c = \frac{1}{\sqrt{\rho\kappa}}$ (where κ is a constant related to the elasticity of the medium). Thus, $Z = \sqrt{\frac{\rho}{\kappa}}$ and c) is correct.

3. (2) If the tension of a vibrating string is increased, the vibration frequency

   a) goes up **X**
   b) stays the same
   c) goes down

   Many ways you could show this, but fundamentally $c = \sqrt{\frac{T}{\epsilon}}$, where c is the wave velocity, T is the tension, and ϵ is the mass density. Since the resonant wavelength is determined by the string length, increasing the velocity also increases the frequency (since $c = freq \times wavelength$). Thus a) is correct.

4. (2) If the mass density in a uniform acoustic tube decreases, its resonant frequencies

   a) go up **X**

   b) stay the same

   c) go down

   A decrease in mass density will cause the volume velocity to increase (by Newton's second law). Since $c = freq \times wavelength$**, a) is correct.**

5. (2) The spectrum of a trumpet note typically

   a) widens **X**

   b) narrows

   c) stays the same

   as the excitation intensity increases.

   a) is correct, perhaps because of the non-linearity at the mouth-piece.

6. (2) Most of the sound energy of a piano comes from the radiation pattern of the struck string (true or false)

   False; it comes from the sound board (which is, of course, mechanically coupled to the string.

7. (8) Describe the sequence of events leading to auditory nerve spikes when an acoustic pressure wave appears at the outer ear.

   Acoustic pressure on the eardrum causes the eardrum to vibrate; this vibration is transmitted via the ossicles in the middle ear. The stapes impinges on the oval window of the inner ear, setting the cochlear fluid into motion. The waves of this fluid cause the basilar membrane to vibrate. Specialized hair cells now begin their transduction. First, shearing of the stereocilia (hairs) opens potassium channels in the hair cells. Since potassium ions are positively charged, hair cell voltage increases (depolarization). This voltage increase opens calcium channels near the base of the hair cell, which in turn releases a neurotransmitter, and this causes the opening of sodium channels in the inervating neural fibers. Depolarization and action potential (a spike) in the fibre is the end result.

8. (2) The bundle of neurons leaving the inner ear contain most of the neurons in the brain's auditory system (true or false)

   False. For example, the auditory cortex is estimated to have ten million neurons, compared to thirty thousand peripheral nerve fibers.

9. (2) The Voder was

   a) a physical model of the human vocal apparatus

   b) an early example of subtractive synthesis **X**

   c) an early example of additive synthesis

   d) a member of the electorate with a head cold

   b) is correct because the Voder produced speech via wide band excitation of a filter bank; there was no explicit model of a tongue, jaw, etc. . d) is an acceptable answer, but only if b) is checked also.

10. (4) During World War II, Roosevelt and Churchill conversed by telephone between London and Washington, using a Vocoder. Explain why the vocoder was an important component of the communications link.

    Superimposing security requirements on a transatlantic link meant increasing the transmission bandwidth. Also, the best encryption could be obtained by first digitizing the speech and this also increased bandwidth. Thus, it was necessary to code the speech with a parametric system of reduced bandwidth, so that additional bandwidth required could be added without exceeding transmission constraints.

11. (6) The human vocal tract has been represented as a time variable filter excited by one or more sources. Discuss the excitation mechanisms for the production of vowels, voiceless fricatives (e.g., s, sh, f, th), and the voiceless plosives (p, t, and k).

    Air pressure is supplied by the lungs. For vowel production, the cyclic opening and closing of the glottis creates a sequence of pressure pulses that excite resonant modes of the vocal tract and nasal tract; the energy created is radiated from the mouth and nose to the listener.

    For voiceless fricatives, the vocal folds are kept open and the air stream is forced through a narrow orifice in the vocal tract to produce a turbulent, noise-like excitation. For example, the constriction for 'th' is between tongue and teeth, and for 'f' is between lips and teeth.

    For voiceless plosives, there is a cross-section of complete closure in the vocal tract, thus causing a pressure buildup. The sudden release creates a transient burst followed by a lengthier period of aspiration.

12. (2) What is a phoneme? (A description of its characteristics or function will be acceptable in lieu of a strict definition).

    Phonemes are "the smallest segments of sounds that can be distinguished by their contrast within words" (Ladefoged). Note that they are linguistic, and not acoustic units, so that phonemes can generally be realized by a number of different sounds.

13. (2) Speech recognition research in the 1950's primarily focused on

    a) the alphabet

    b) isolated digits **A**

    c) toy dogs

    d) continuous speech

    (Mention an example).

    Audrey was an isolated digits recognizer designed and built at Bell Labs in the 1950's. Like nearly all published recognizers then and since, its accuracy was in the high 90's.

14. (2) Roughly what is the wavelength of a 1 kHz sound wave travelling in air at room temperature?

    Roughly a foot. Remember the "foot per millisecond" rule for the speed of sound? Just divide by the frequency ...

15. (6) List two different mechanisms for a human listener to locate a sound source laterally in space (that is, to the left or right). Answer from the functional or engineering perspective (that is, don't worry about the physiology per se). What portion of the acoustic signal is used primarily by each mechanism?

    The first mechanism is time delay or phase difference between the acoustic wave's arrival at the two ears. For the phase difference to be unambiguous, the phase difference between the two ears should be significantly less than a cycle. Thus the part of the acoustic signal that this works well for (in humans) is low-mid frequencies (below a few kHz).

    The second mechanism is amplitude difference, which is minimal for low frequencies - the head must act as a significant barrier for there to be significant amplitude differences at the ear. Thus, the part of the acoustic signal that this works well for (in humans) is high frequencies (above 6 kHz), for which the wavelength is significantly smaller than the head.

    In the middle range of frequencies (say 2-6 kHz), both mechanisms operate to some extent.

16. (2) Having a long ($>$ 2 sec) reverberation time for a room is typically

    a) good for

    b) bad for **X**

    c) irrelevant for

    the intelligibility of speech.

    Definitely bad! Long reverberation times temporally smear the speech across phone and word boundaries, so that low-amplitude phones can sometimes be masked.

17. (2) LPC and DTW were

    a) Major developments in speech technology in the 60's and 70's **X**

    b) Major developments in speech technology in the 80's

    c) The initials of two commentators on ABC's Wide World of Sports

    (Expand the acronyms).

    DTW - Dynamic Time Warping - independently developed by several people in Japan and Europe in the late 1960's; further developed and applied in the early to mid 1970's.

    LPC - Linear Predictive Coding - independently developed by several people in Japan, U.S., and Europe in the late 1960's; further developed, applied, and properly explained in a number of key papers and books in the 1970's.

18. (5) Associate 5 of these names (you can skip one) with developments in science or engineering. **Here are *some* reasonable answers; others are possible.**

    - Dudley - **Voder and Vocoder**
    - Helmholtz - **Filter bank theory of auditory perception**
    - von Kempelen - **First basically complete mechanical speech synthesizer, plus a detailed book on his work.**
    - Wheatstone - **Mechanical speech synthesizer based on von Kempelen's.**
    - von Bekesy - **First measurement of human basilar membrane motion, partly verifying Helmholtz proposal.**
    - Kiang - **pioneered electrical measurement methods for single neurons.**

19. (5) Choose an engineering application in speech, music, or general audio, and explain how in principle the study of some aspect of human performance could potentially be (or is already) helpful for this application. Mention some specific characteristics of the human example.

    Again, obviously there were many possible reasonable answers to this question. Robustness to the vocal tract size of a speaker as explored by PLP could be one good direction for an answer. Here is another:

    Application: speech recognition. Humans performance at this task is relatively robust to noise, perhaps partially due to some properties of the peripheral auditory system. If we can derive engineering models that in principle have some of these properties, we can make some robust systems. For example, Tom Holton's models could potentially lead to more robust formant detection, which in turn could be useful for speech recognition. These models appear to be quite robust in noise, since they rely on phase coherence that is quite local, as well as on a strong model for the nature of voiced speech (impulse followed by synchronous region). His model is based not only on this characteristic of voiced speech, but also on a model of the auditory periphery. If we could learn more about the accuracy or inaccuracy of this model, we could get a better idea about the only speech recognition system we know that has such a strong robustness to noise, and this knowledge could be a strong guide in our future experiments and designs.