Under what assumptions does sampling capture all of the information in a signal?

1. What determines the range of frequencies you can capture?

half of the sampling rate.

1. What determines the signal-to-noise ratio of the captured signal?

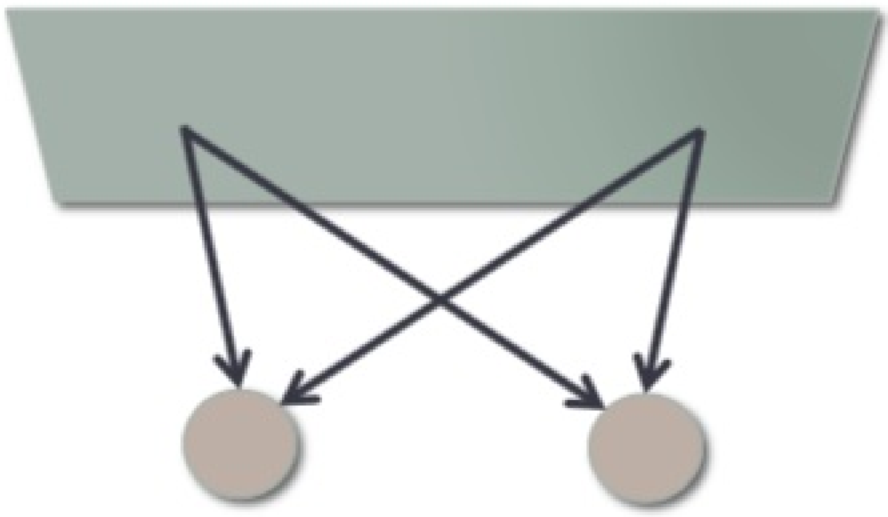
bits per sample.

1. You should know how to represent sinusoids of varying amplitudes and frequencies in a spectrum (frequency domain) plot.

2. You should know how to interpret a filter frequency response plot, e.g. draw frequency responses for low-pass, band-pass, and high-pass filters. Show the effect of applying a given filter to a given signal using frequency domain representations.

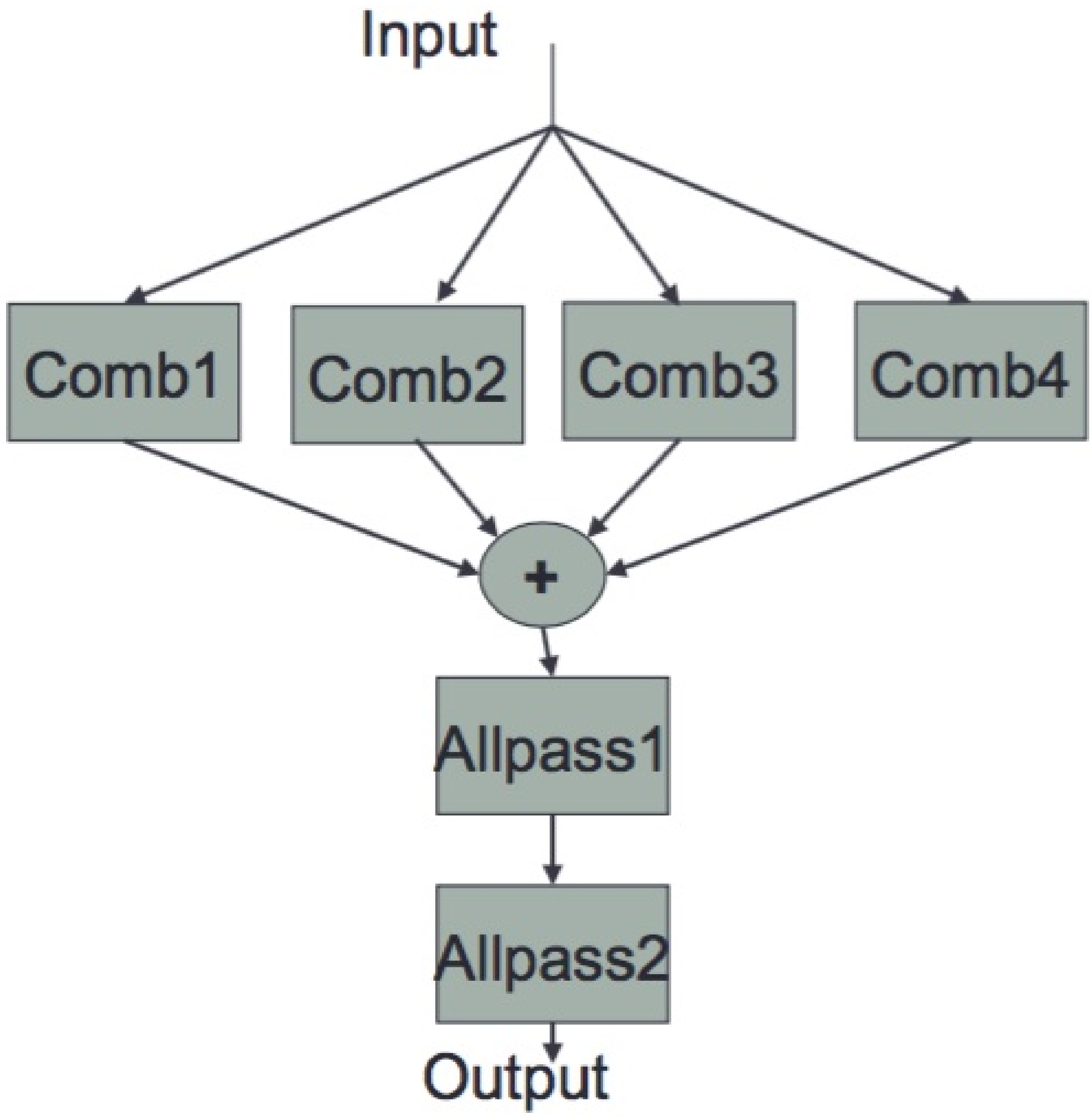
3. If you wanted to create reverberation to match that of Carnegie Hall using convolution, what would you do? Describe what measurement you would make in the hall, what sound source and where, where to put microphones, what is the recording called, how do you use it to simulate reverberation? (We're not looking for detailed audio engineering, but conceptually, how does this work?)

measure impulse response of the hall. Get four impulse responses for stereo.Multiply impulse responses using fast convolution. Reverberators can be seen as big filters. I will use the result to build models for the hall.



What are the main components of Schroeder-style reverb?

4 comb filters and 2 all pass filters.



What is RT60? (relating to reverberation)

decay time for peak amplitude to decay to -60dB(1/1000 amplitude) usually 1.5-3 seconds. decay exponentially.

Of the two main types of reverb, Schroeder-style and Convolution, which gives the most accurate model of a room? Which gives the most parametric control? Which is computationally the most efficient?

convolution is more accurate, the most parametric control, the most efficient.

You should know about amplitude, power, and intensity. What change in power corresponds to 10dB? What change in amplitude corresponds to 10dB? What change in amplitude corresponds to 6dB? (We did not talk about intensity much in class – intensity is power per unit of area, so at a fixed distance from a speaker, intensity is proportional to power.)

10 times. root of 10 times. root of 2 times.

How does intensity decrease with distance (assuming no reflections or absorption)? How does amplitude decrease with distance?

doubling the distance, -6dB. the same.

What is the speed of sound?

1ft/ms

What is the smallest difference in amplitude we can hear?

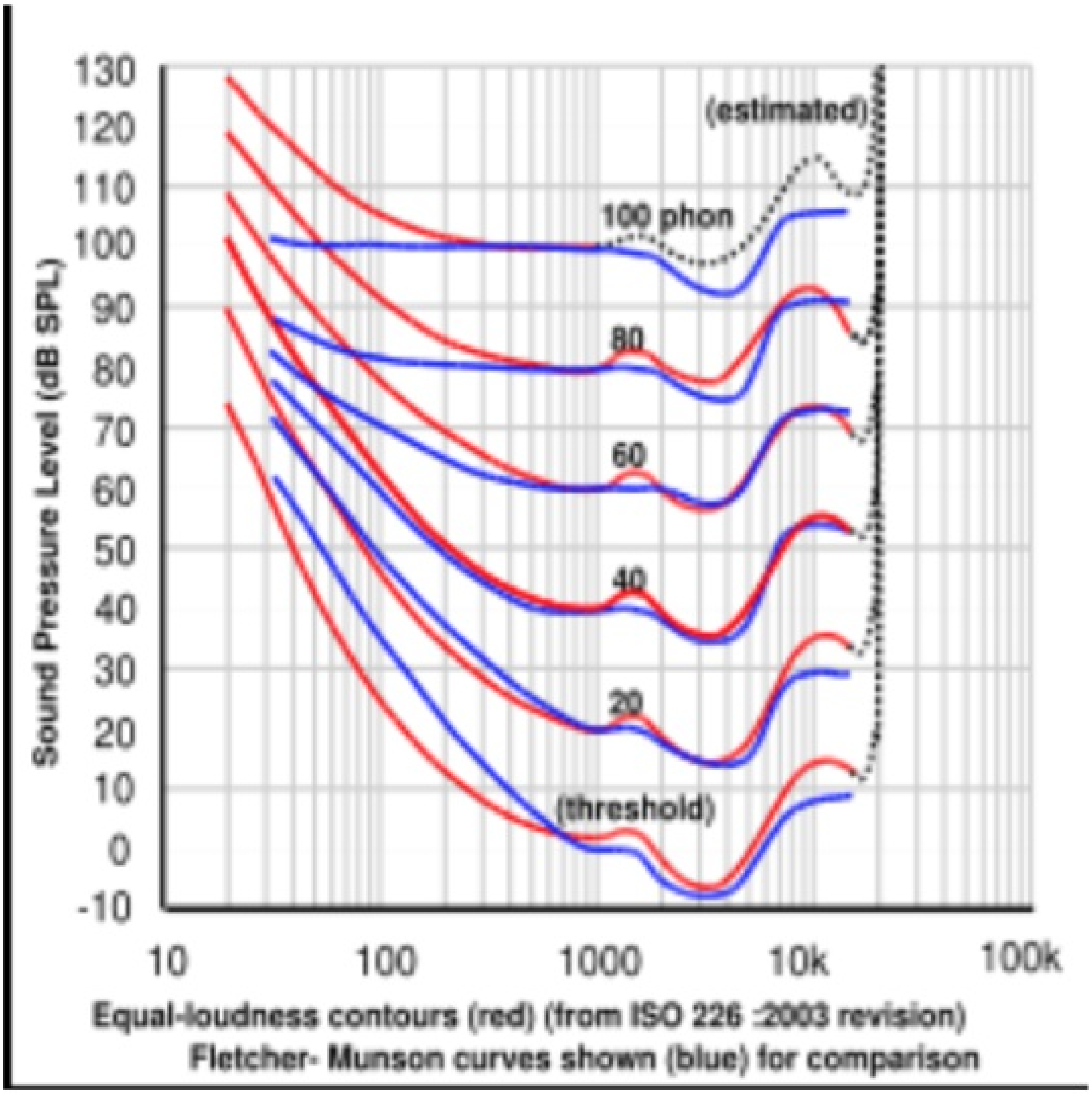
1.12m

What is the smallest difference in frequency we can hear?

0.7567Hz

Intensity I is the energy E flowing across a unit surface area per unit of time t. the intensity range of human hearing is 120 dB.

Explain the Fletcher Munson graph. How much louder is the threshold of hearing at 100 Hz than at 1000 Hz? If you change the frequency of a sinusoid at 100Hz and 70 dB SPL to 1000Hz, how many dB quieter should you make the sound to give the same perceptual loudness?



people perception of loudness is more sensitive as frequency rises. 10 times louder. 10dB。

the left speaker is at 0 radians and the right speaker is at π radians, giving us a panning range of $\theta \in [0; \pi]$, with the center position at $\theta = \pi$. linear panning: $L(\theta) + R(\theta) = 1$. 耳朵垂直线和左边的夹角是θ，linear的算法是，左边（pi/2 - θ) * 2/pi 右边 θ ＊ 2 ／ pi

1. What is constant power panning? Why would you use constant power panning? With constant power panning, if a sound has an amplitude of 1 when panned full left, what amplitude do you send to both speakers when panning to the center? $L(\theta) = \cos(\theta)$ and $R(\theta) = \sin(\theta)$ One way of dealing with the "hole-in-the-middle" effect is to

use constant power panning. Basically, we exchange the linear function for L(θ) and R(θ) with the sine and cosine functions, such that L(θ) = cos(θ) and R(θ) = sin(θ). Main idea: power is proportional to the squared amplitude, and cos2 + sin2 = 1, thus L(θ) = cos(θ) and R(θ) = sin(θ) yields constant power panning. pan full left angle = 0

In such a case, if the listener is always placed at the sweet-spot, and facing the speakers, the hole-in-the-middle effect may not be very significant. However, if the speakers are placed far from each other in a big room, then the phases

1. What are the three types of data compression? Define these terms and give an example of each. coding redundancy: variable-length coding （pcm,ulaw,alaw) compress based on inter-sample redundancy: zip file\midi file compress based on psycho-perceptual redundancy: jpeg

2. What is frequency masking? Explain the roll of frequency masking in MP3 music compression. Using ear as filter bank. loud sound in a channel masks soft sound in adjacent channels. hide equalization effects.

3. You should be familiar with the steps of MP3 encoding. E.g. what is the purpose of Huffman Coding in MP3?1 Why does MP3 encode music using a frequency-based representation rather than a time-based one? step 1: filter bank(32 bands of equal width) step 2: psychoacoustic model(Take 1024 point FFT of audio, group spectrum into critical bands;Identify tonal components (sinusoids) and assign tonal index to each critical band) huffman: removing coding redundancy: Gives smallest number of code symbols per source symbol. this allows more efficient coding. apply filters

because in time domain we can't view frequency directly. it's good to transfer continuous data into discrete data.

4. Suppose you are concerned about quantization noise using an 8-bit DAC. What is the improvement in S/N ratio if you upgrade to a 12-bit DAC (assume perfect, noise-free analog circuits)? 4 * 6 = 24dB

5. Why do many telephone systems use (8-bit) µLaw rather than 8-bit PCM? Explain an advantage of µLaw over PCM and an advantage of PCM over µLaw. a. scaled to span 13 bit dynamic range. a large range for representation. b. because ulaw is logarithmic. there are more codes for near-zero amplitudes than high amplitude, while pcm is linear, more equally distributed.

6. Roughly how many data bits are on an audio CD (assume 80 minutes)? Hint: 16-bit stereo at 44100 Hz. 16 *44100* 2 * 4800

   Use Nyquist patterns to generate these (infinite) sequences:

   123123123...

   112123123412345 1121231234 12345 ... 123321123321...

make-cycle({a b c}) abcabc make-panlindrome({a b c}, elide: first) abcbaabcba make-heap({a b c}, max: 1) abccabcbaacb(every element randomly chosen) make-accumulation({a b c}) aababcaababc make-copier(make-cycle({a b c}, for: 1), repeat: 2) aabbcc make-accumulate(make-cycle({1 3 -4})) 1 4 0 make-line({a b c}) abcccccc

Write a Nyquist function to produce an FM tone with approximately N harmonics starting at F Hz and with amplitude A, where N, F, and A, are parameters.

fmosc(hz-to-step(F), hzosc(F) *(N - 2)* F)

What is computer accompaniment?

You should be familiar with the topics in the music understanding lecture. E.g. define the following terms: style recognition, onset detection, score alignment, music fingerprinting, music recommendation, music distribution, human computer music performance (of popular music).

Multiplication in the time domain is equivalent to what in the frequency domain?

Multiplication in the frequency domain is equivalent to what in the amplitude domain?

Multiplying by a constant in the time domain is equivalent to what in the frequency domain? (Careful: you might think the answer has to do with convolution, but this is a special case with an especially simple answer.)

Adding two functions in the time domain is equivalent to what with the functions in the frequency domain?

Superposition principles

in a linear system, different filter can be done through simple sum.

How would you accomplish time stretching with granular synthesis?

Given a time-stretching method, how would you accomplish pitch shifting (without stretching or shrinking)?

How does sampling synthesis achieve control of duration and pitch?

How does sampling avoid the problem that pitch-shifting a recording can sound unnatural if you pitch shift by a large amount?

do multi pitch range sampling and avoid large amount shifting.

What are the formant frequencies for the "oo (who)" and "ee" vowels?

You should understand the components of physical models such as the clarinet and bowed string models discussed in class. E.g. identify the waveguide, lumped filter, and non-linear elements in the clarinet and string models. What do these correspond to in corresponding acoustic instruments?

mcm models

Delay-line loop of one period • Low-pass filter modeling losses over one loop • Non-linear element to generate oscillation

How does Karplus Strong synthesis work? Describe the elements and what they do? E.g. how is pitch determined, and what causes the string to decay?

Fill table with noise or initial conditions • Perform table-lookup oscillator on noise • Phase-increment = 1 • Average adjacent samples as they are read

What is commuted synthesis?

drive the string with impulse response of body When bow slips on string, it generates a sort of impulse At every bow slip, insert body impulse response into string model

You should understand the concepts of spectral interpolation synthesis, e.g. how are spectra selected?

use two tables. Interpolate between tables Keep phases coherent so that interpolation is truly interpolation of spectra only choose harmonic spectra.

Why are there two tables in the synthesizer? What goes into the

tables?

What are HRTFs?

head related transfer function. a response that characterizes how an ear receives a sound from a point in space;

How can we perceive elevation or amplitude when the duplex theory predicts a cone of confusion? Why are HRTFs usually used with headphones?

distance has close relationship with amplitude and elevation.

How would you simulate spatial location in a concert hall?

build model for the hall. using responses for impulses to calculate the amplitude.

John Chowning made the point that things do not necessarily seem more distant just because you lower the amplitude. What cues are available for distance?

1. In terms of what is sound intensity measured? Write the expression relating this measure with pressure. w/m^2 p/a
2. What is the most sensitive frequency region of the human ear? 2000-5000hz

3. In three sentences, summarize the functionalities of the outer ear, middle ear and the inner ear. gather sound wave. amplitude frequency timbre

4. What is the difference between reverberation and echo? A reverberation is perceived when the reflected sound wave reaches your ear in less than 0.1 second after the original sound wave. Since the original sound wave is still held in memory, there is no time delay

between the perception of the reflected sound wave and the original sound wave.

5. If two sine tones are very close together in frequency, the total loudness perceived is less than the sum of the two loudnesses that would be heard from the tones played separately. Why does this happen?

6. Why is white noise called "white"?

7. What happens if a set of partials is modulated in frequency by a common temporal envelope? they will share the change in frequency.

8. What is the missing-fundamental phenomenon?

LPC = Linear Prediction Coding • Model: predict next sample as a weighted sum of past samples.

1. What are the two aspects of Lord Rayleigh's Duplex Theory of sound localization? a. Interaural Time Difference and Interaural Level Difference b. Interaural Time Difference and Interaural Frequency Difference c. Interaural Transfer Difference and Interaural Frequency Difference d. Interaural Transfer Difference and Interaural Elevation Difference Answer: ___
2. Which of the following has the greatest effect in forming the Head Related Transfer Function (HRTF)? a. Eyes b. Ears c. Hair d. Mouth Answer: ___
3. Which of the following is NOT a technique originally developed for voice synthesis? a. Linear Prediction Coding (LPC) b. VOSIM c. Formant W ave-Function Synthesis (FOF) d. Digital Pulse Code Modulation (DPCM) Answer:

---

5 Study Guide: 15-322 Introduction to Computer Music

1. In Linear Panning between 2 speakers, a value α, between 0 and 1, is the portion of the sound coming from the left speaker, and (1- α) is the portion of the sound coming from the right speaker. What is the main problem with using Linear Panning to make a sound “move” from left to right (or vice versa)? a. Mid-range frequencies are attenuated b. It’ s electronically hard to implement c. The power of the signal changes with α, so the perceived loudness of the sound varies with pan position d. Instead of moving from left to right (or vice versa), the sound will seem to move from above to below (or vice versa) Answer: ___
2. The advent of “low-fi” popular music in the eighties was brought about by a flood of reasonable-quality, low-cost, a. Magnetic tape (brought home from Germany after the war, but only available cheaply after about 1978) b. Home analog multitrack recording machines c. CD-ROM – equipped “multimedia” PC’s d. Microphones, which replaced the brittle crystal diaphragms of older studios Answer: ___
3. What is the hardware arrangement for sound delivery, given a goal of creating the best illusion of 3-D sound using the HRTF? a. Headphones b. Small speakers a few feet away from the ears, such as those of a desktop computer. c. Large wall-mounted stereo speakers. d. 5 or 6 surround / wraparound speakers, such as a Dolby 5.1 arrangement in a room. Answer: ___
4. What defines the human range of hearing (typically), and about how large is it? a. quiet whisper – loud shout, about 60 dB b. softest perceivable sound – loudest painless (or softest painful) sound, 120 dB c. eardrum resonance – eardrum rupture, about 200 dB d. there is no limit to human hearing in either loudness or softness, the only reason some noises escape our perception is because they are

below the noise threshold of the air around us. 6 Study Guide: 15-322 Introduction to Computer Music Answer: ___

5. The fact that piano wires are made out of steel, which has some inherent rigidity means that the harmonics of a vibrating piano string (in terms of the fundamental frequency f) are actually about a. f, 2.001*f, 3.004*f, 4.010 *f... b. f, 1.999* f, 2.996 *f, 3.090* f... c. f, 1.5 *f, 2.25* f, (1.5 cubed) * f... d. The vibration is an interaction between gravity and wire tension, and material properties play no measureable part. Answer: ___
6. There are five formants in the vowel sounds produced by the human voice, but the most important ones that human listeners use to understand one another are a. The bottom two (in frequency) b. The top two (in frequency) c. The second and third (in frequency, from lowest to highest) d. The highest and lowest (in frequency). Answer: ___
7. An impulse response is: a. The signal output by a given filter when the input is just an impulse b. A filter output that consists of just an impulse c. The fastest output a filter can generate d. A certain kind of filter Answer: ___
8. Fill in the blank: __ in the time domain is equivalent to multiplication in the frequency domain. a. Addition b. Superposition c. Amplitude modulation d. Convolution Answer: __
9. Fill in the blank: _ in the frequency domain is equivalent to multiplication in the time domain. 7 Study Guide: 15-322 Introduction to Computer Music a. Addition b. Superposition c. Amplitude modulation d. Convolution Answer: __
10. Here is the frequency response for a given filter (only magnitude is shown): Amplitude What kind of filter is this? a. Low-pass b. High-pass c. All-pass d. Forward-pass Answer: __ Frequency
11. Which of the following phenomena is responsible for the behavior of a comb filter? a. Low-pass filtering b. High-pass filtering c.

Constructive interference d. Superposition Answer: __

12. A circuit (see figure) is constructed such that its output is defined recursively by: Out(t + 0.0025 seconds) = In(t) + 0.5 * Out(t). In(t) a. Low-pass filter with the -3 dB point at 250 Hz b. Low-pass filter with the -3 dB point at 400 Hz c. Comb filter strengthening integer multiples (harmonics) of 250 Hz d. Comb filter strengthening integer multiples (harmonics) of 400 Hz Answer: __
    - Out(t) This circuit behaves as a: 0.0025s delay 8 Study Guide: 15-322 Introduction to Computer Music
13. From an original waveform a frequency spectrum is generated by means of a Short-Time Fourier Transform. Were a new waveform to be generated from that spectrum by some inverse operation of the Short-Time Fourier Transform, the new waveform would a. Compare favorably with the original waveform b. Be recognizable as a highly distorted version of the original waveform c. Be completely unrecognizable noise d. The Short-Term Fourier Transform is a mathematical “one-way” function, there is no way to invert it and go from spectra to waveforms. Answer: __
14. Which of the following is not a cue for judging how far away a sound was generated? a. Diminished amplitude / diminished energy b. Shifted phase c. Loss of high-frequency components d. Loss of low-amplitude frequencies below noise or hearing threshold. Answer: __
15. Which cue is most indicative of a sound being generated by a source in motion? a. Increasing amplitude b. Decreasing amplitude c. Doppler-shifted frequencies d. Frequency-variable phase shifts Answer: __
16. Which of the following will the sound generated by the Nyquist code (play (fmosc c5 (mult (pwl 0.05 2000 1 1000 1) (osc c5)))) sound most like? a. A click b. Hitting a metallic pipe c. A singer’s voice d. A brass instrument Answer: _ 9