Quiz 1: CSL 607 Multimedia Systems

Duration: 1 hour Total Marks: 50

Name: Roll Number:

1 State true or false with explanation. Explanation should not be more than three lines. 15x2=30 Marks

1. One Huffman code is prefix of other Huffman code, therefore they are also called prefix codes.
   Ans: False. One Huffman code cannot be prefix of other Huffman code. It is actually true for any variable length coding. This is required for the decoder to interpret the variable length codes. As soon as decoder receives a valid symbol, it spits out the corresponding symbol.

2. For the analog to digital conversion, the input signal is directly sent to the sampler without any pre-processing.
   Ans: False. The signal is first sent to a low-pass filter in order to avoid aliasing.

3. Loudness is linearly proportional to the audio intensity.
   Ans: False. Loudness and intensity are non-linearly related. Loudness is roughly proportional to the logarithmic of the intensity.

4. Pitch can be perceived only when the fundamental frequency is present.
   Ans: False. Pitch can be perceived even without fundamental frequency component. Human auditory system makes up for the missing fundamental component.

5. Human auditory system functions as a transducer to convert energy from one form to another form. Ans: True. Human auditory system converts pressure waves into electrical impulses.

6. Masking phenomenon happens both in time and frequency domain.
   Ans: True. In frequency-domain masking, a loud tone can mask higher tones. Similarly, in temporal masking a loud tone saturates the hearing receptors in the inner ear which take time to recover.

7. To improve the speech quality in the VOIP (voice over internet protocol), it is good idea to raise the frequency to 44.1 kHz because the sound quality is proportional to the sampling frequency.
   Ans: False. Voice does not have frequencies above 4kHz, so 8kHz sampling frequency is enough.

8. For short-term frequency analysis of audio signal, it is good idea to use rectangular window because it preserves original sample amplitudes.
   Ans: False. Rectangular time window distorts the frequency response. The rectangular window causes the spectrum to be spread across other frequencies as well which are not present in the signal. The spectral spread is less with a Hamming window.

9. A larger number of filter coefficients ensures smaller transition band, but need more computing power.
   Ans: True. With more filter coefficients, the spectral spread is less and the frequency response is close to the desired frequency response.

10. If we split the entire frequency band of music into two non-overlapping and equal frequency bands, the perceptual relevance of the lower frequency band is more than higher band .
    Ans: True. This is because of the structure of the human auditory system. At higher frequencies, humans ability to perceive frequency change reduces.

11. In Fourier analysis, we usually illustrate the amplitude spectrum, because the phase information irrelevant for sound synthesis and human perception.
Ans: True. Generally, the relative phase distortion has no audible effect. Yet, many research papers have found the phase distortion perceivable in certain conditions.

12. We cannot differentiate two music instrument sounds that are similar in both pitch and loudness.
Ans: False. The two instruments will have different timbre which helps us distinguish two instruments. Timbre is mainly determined by the harmonic content of the sound. It takes some time ($\approx$ 50ms) for us to perceive timbre. Therefore, it is hard to distinguish very short tones from different instruments.

13. Because pitch is measured in terms of fundamental frequency of the speech signal, fundamental frequency is necessary to perceive pitch. Ans: False

14. Gaussian Mixture Model assumes that, for a given speaker, features distribution has a single peak.
Ans: False: There can be multiple peaks, each modeled by a Gaussian.

15. In vector quantization, the codebook size increases linearly with the number of feature vectors.
Ans: False: The codebook size is fixed. When more feature vectors are added, the codes may change, but the codebook size remains constant.

2 Short answer questions, maximum 5 lines. 5x4=20 Marks

1. The maximum audio sampling frequency that you can have is fixed, say 24 K samples per second. How would you ensure no aliasing?
Ans: Low pass filtering before sampling.

2. How are temporal and frequency masking phenomenon used for audio compression?
Ans: As per lecture.

3. What are audio fingerprints? Write three characteristics of audio fingerprints?
Ans: A compact vector representing an audio segment. Compact, discriminatory, efficient to calculate.

4. To compress audio signals effectively, what are the two fundamental methodologies that can be used in perceptual audio codec?
Ans: Quantization, masking, entropy coding, run length coding.

5. Why do we calculate Modified Discrete Cosine Transform (MDCT) of the filter-bank outputs in MP3 codec?
Ans: To exploit correlation among filter band outputs.