Example answer to question 1:

a) *Symbolic representation* (e.g. KERN, MIDI, MusicXML): in this representation, musical content is encoded through time-stamped events, such as pitch, onset, and offset for encoding a musical note event. *Digital audio representation (e.g. MP3, Wav)*: this representation is derived directly from the sound signal, allowing the extraction of low-level features such as fundamental frequency, amplitude envelope, attack.

b) *Symbolic representation*: in order to capture the fine-grained similarities between related folk songs, melodic similarity between these songs is often determined based on a symbolic representation of melodies. An application scenario of symbolic similarity measures for melodies are search engines in cultural heritage institutions which allow the general public to access and browse through the collection online. *Digital audio representation*: digital audio is very suited for investigations into psycho-acoustic features such as loudness, sharpness and roughness, or timbral features. Applications of these features are corpus–based studies that investigate, for instance, what distinguishes a chorus from a verse in a pop song using these features.

Example answer to question 2:

Answer to 2a:

- Melodic information: information extracted from melodies can be used to compute the similarity between songs. For instance, extracting information on pitch and duration for each note of the melody allows to compute similarity between songs using the Earth Mover's Distance.
- Rhythmic information: information extracted from musical pieces on rhythmic patterns allows to compute rhythmic similarity in order to distinguish different dance styles. One possible way is to extract all onsets of a piece and to compute underlying metric accent structures.
- Chroma-based information: chroma-based features extracted from audio closely relate to aspects of harmony, and can be used to compute self-similarity matrices, allowing to detect repetition of similar segments between and across musical pieces.

Answer to 2b:

- Challenge of modelling subjectivity: the similarity between two different objects depends on many features characterizing these objects. Different individuals can weigh the features in many different ways for concluding an overall sense of similarity. As we experience music on an every day basis in a personal manner, it can be challenging to find one computational model that weighs all different musical features into one similarity measures that fits different people.
- Challenge of modelling flexibility: depending on the context that two objects are placed in, the similarity between these two objects can be experienced differently even by the same individual, as the features may contribute in different ways to the overall similarity. For instance, we can either emphasize more the commonalities or more the differences between the two objects depending on the context. Music is a very complex phenomenon, with many different features present, requiring flexible models on what features contributes how much in a given context to the overall perceived similarity.

Answer to 2c:

- I describe the computation of rhythmic similarity based on the Inner Metric Analysis model. We need a symbolically encoded music file and extract all onset times of the notes. Within the set of all onsets, we search for all local meters, defined by sets of equally distanced onsets that are maximal (not subset of any other set of equally distanced onsets), in order to determine the metric weight profile of the piece. For calculating the metric weight of a specific note, we take into account all local meters that coincide with it, specifically we calculate the weighted sum over the length of all local meters that coincide with the note. For calculating rhythmic similarity between two pieces, we determine the metric weight profiles for both pieces. After normalizing the weight profiles of both pieces, we then calculate the correlation coefficient between the weight profiles as an indication of the rhythmic similarity between these two pieces.

Example answer to question 3:
Answer to 3a:

- Gaps in the musical surface: whenever there appears to be a perceptual gap in the musical surface (e.g. large melodic jump), listeners tend to segment music.
- Repetition: when listeners perceive recurring patterns, they use the patterns to segment music.
- Closure: when listeners perceive the closure for instance of a melodic line, they tend to perceive that as a segment ending.

Answer to 3b:
I choose the Gestalt-based model of LBDM that detects local discontinuities in melodies regarding pitch and rhythm. We first extract for each note in the melody the pitch information, onset information and offset information. We then compute the profiles of consecutive pitch, interonset and interoffset intervals for measuring the amount of change in pitch and rhythm between consecutive notes. We then compute the boundary strength for each interval, taking into account these profiles, and compute the combined boundary strength profile. The local peaks that occur in the boundary strength profile are then the candidates for the boundaries of melodic segments.