Paper title: Using Deep Audio Embeddings for Music Emotion Recognition

The following contents are included in this supplementary material.

- TSNE Visualization of L3-Net embeddings on 4Q Audio Emotion dataset
- TSNE Visualization of Feature sets by Panda et al. on 4Q Audio Emotion dataset
  (Features were extracted by a baseline MER model by Panda et al.)
- TSNE Visualization of VGGish embeddings on 4Q Audio Emotion dataset
- TSNE Visualization of L3-Net embeddings on Bi-modal Emotion dataset
- TSNE Visualization of VGGish embeddings on Bi-modal Emotion dataset
- TSNE Visualization of L3-Net embeddings on Emotion in Music dataset
- TSNE Visualization of VGGish embeddings on Emotion in Music dataset
- TSNE Visualization of L3-Net embeddings on RAVDESS singing dataset
- TSNE Visualization of VGGish embeddings on RAVDESS singing dataset

The following figures are TSNE Visualization of how the L3-Net and VGGish embedding method can classify different emotion categories. Each point denotes an audio clip. The x and y axes denote T-SNE projection axes.
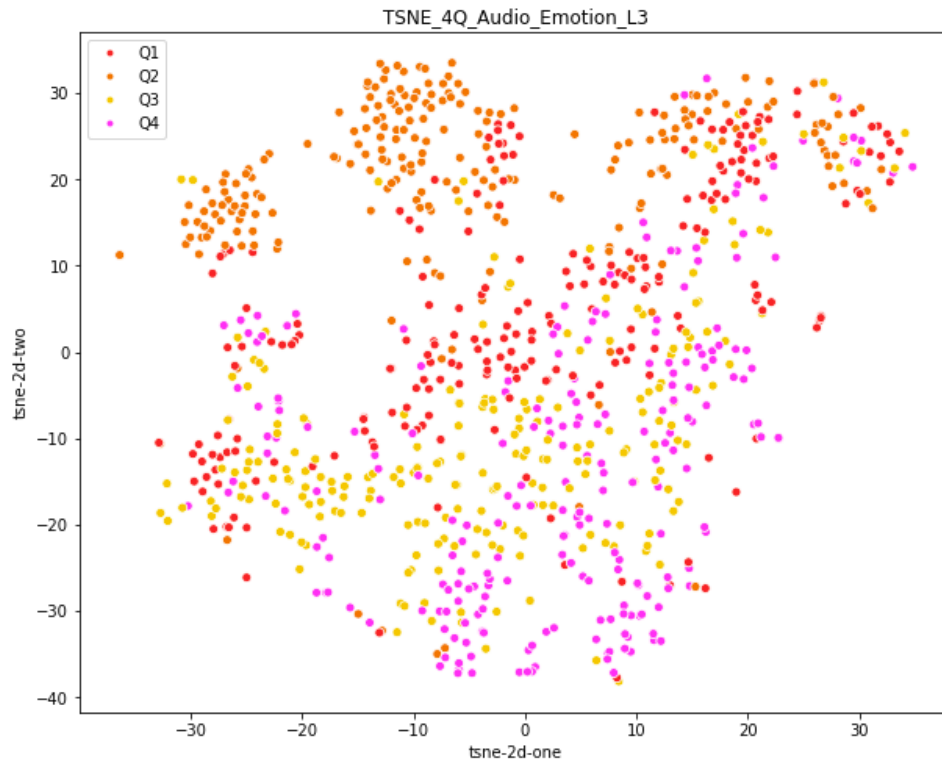

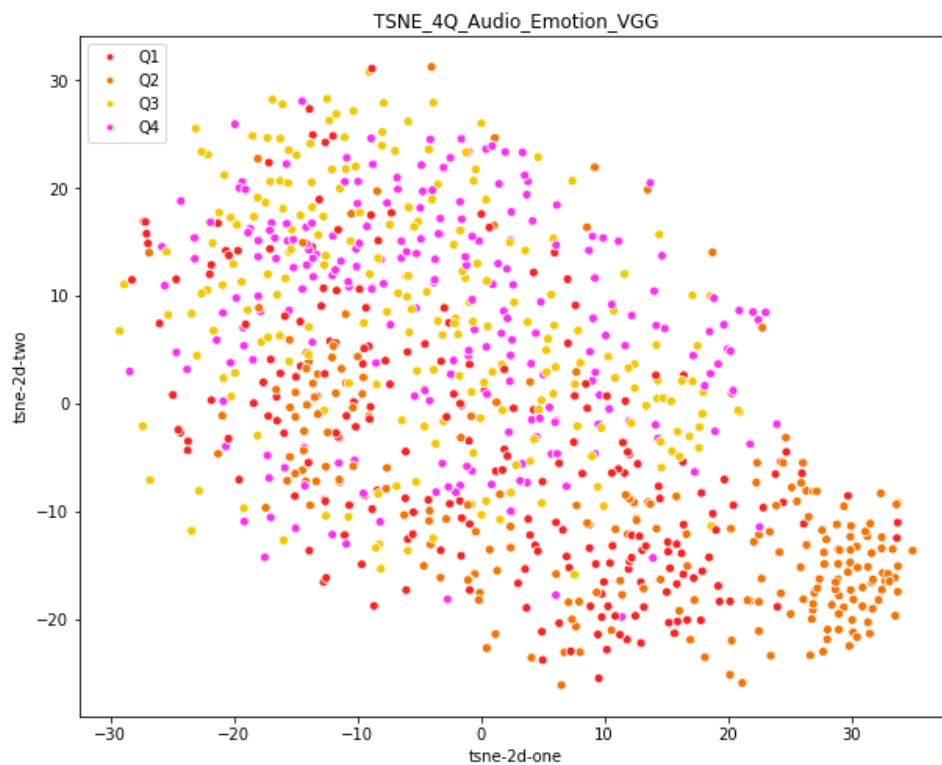Figure 1-1: TSNE of L3-Net embeddings on 4Q Audio Emotion dataset

Figure 1-2: TSNE of VGGish embeddings on 4Q Audio Emotion dataset
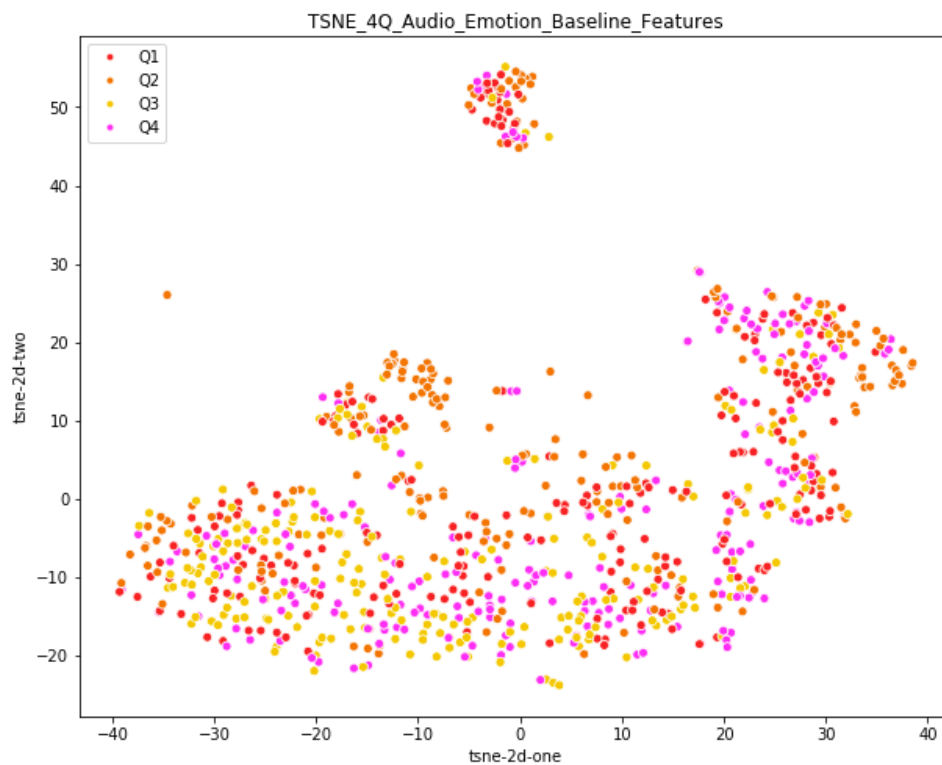

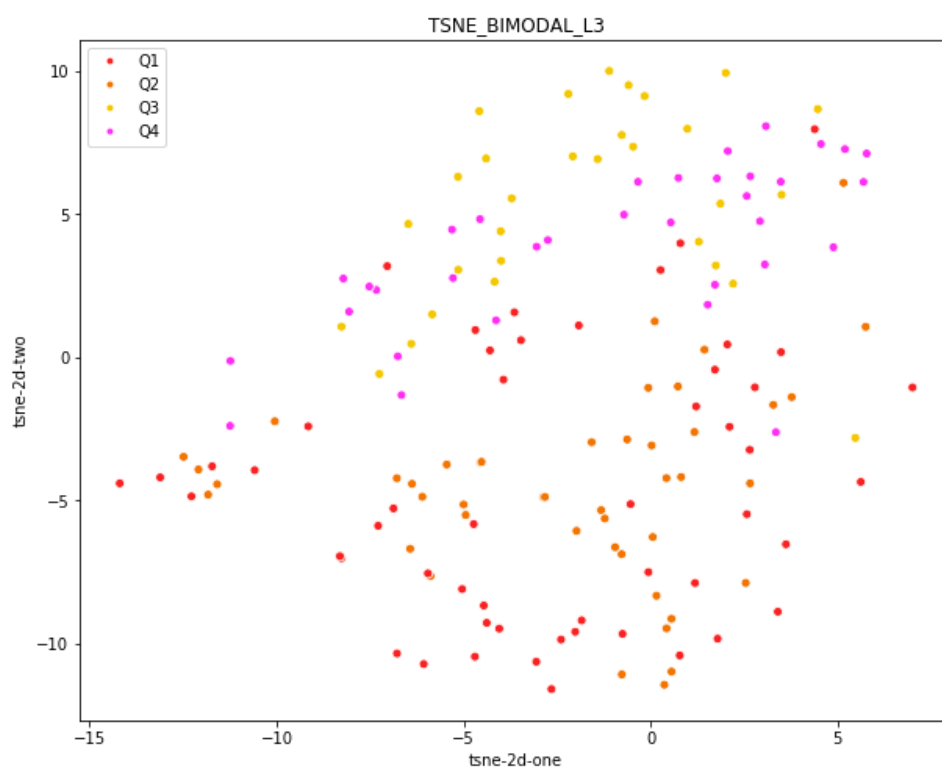Figure 1-3: TSNE of Feature sets by Panda et al. on 4Q Audio Emotion dataset

Figure 1-4: TSNE of L3-Net embeddings on Bi-modal Emotion dataset
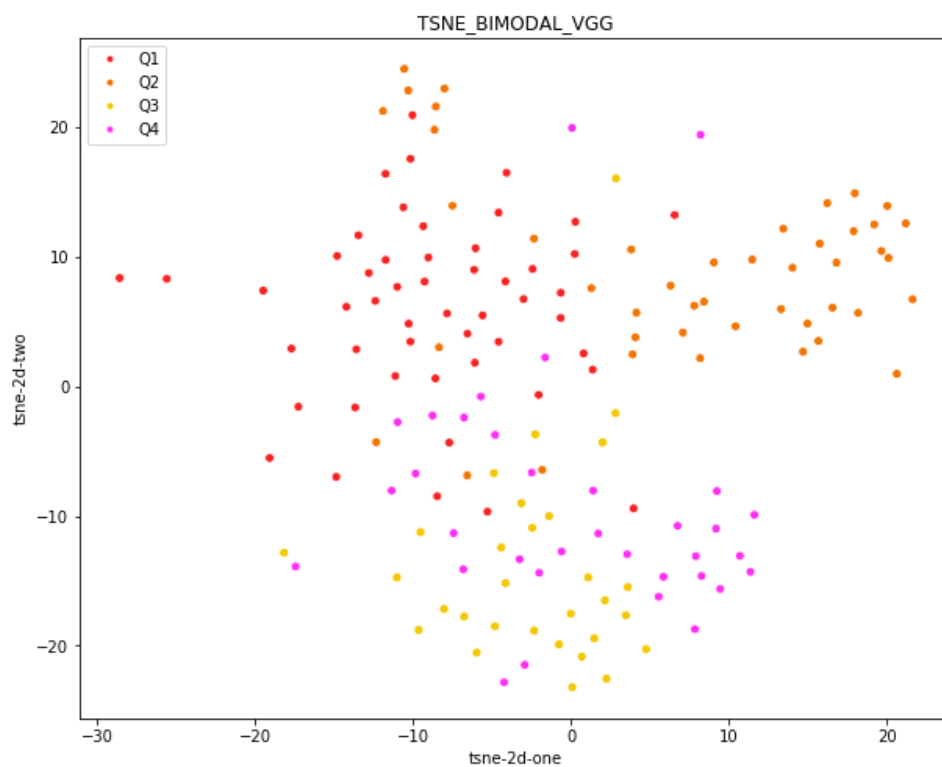

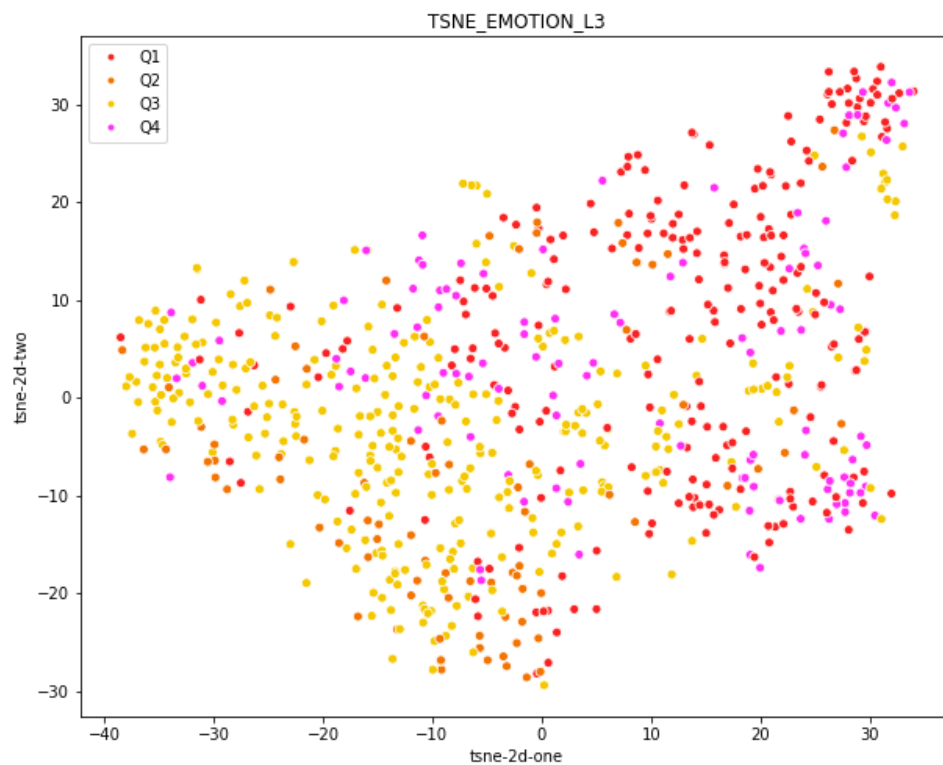Figure 1-5: TSNE of VGGish embeddings on Bi-modal Emotion dataset

Figure 1-6: TSNE of L3-Net embeddings on Emotion in Music dataset
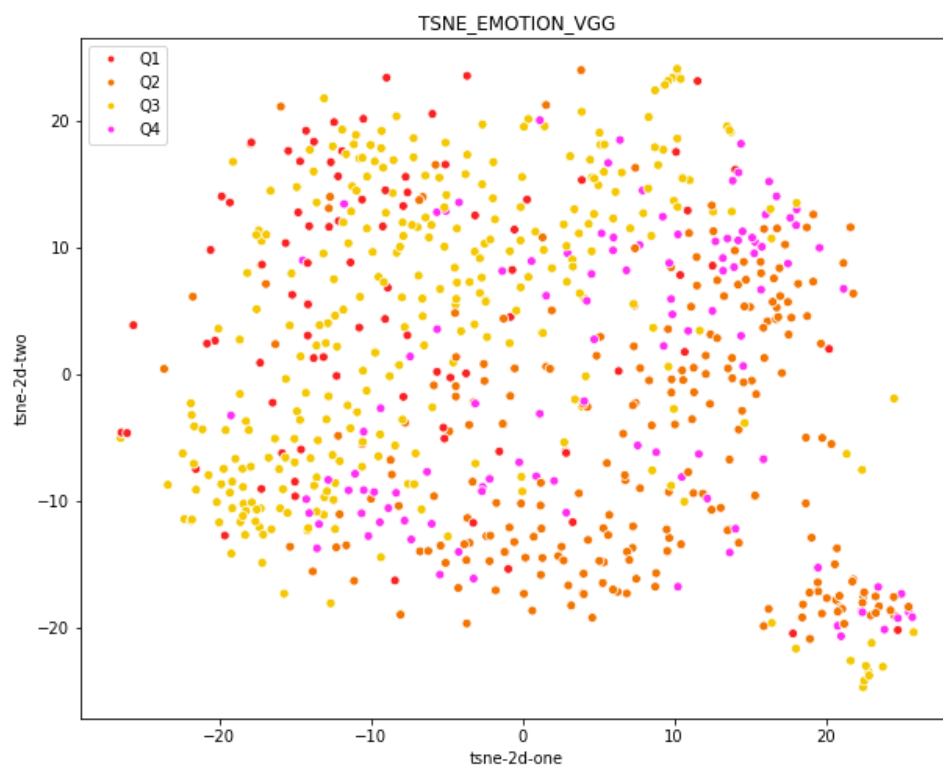

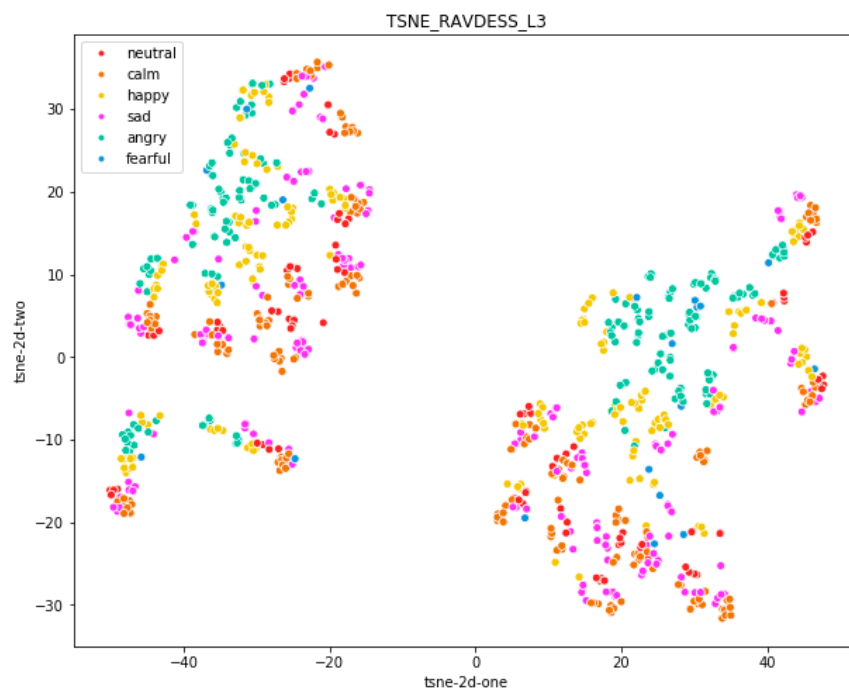Figure 1-7: TSNE of VGGish embeddings on Emotion in Music dataset

Figure 1-8: TSNE of L3-Net embeddings on RAVDESS dataset (clusters are located in the left/right sides of the figure based on female/male singer, and small clusters indicate each singer identification)
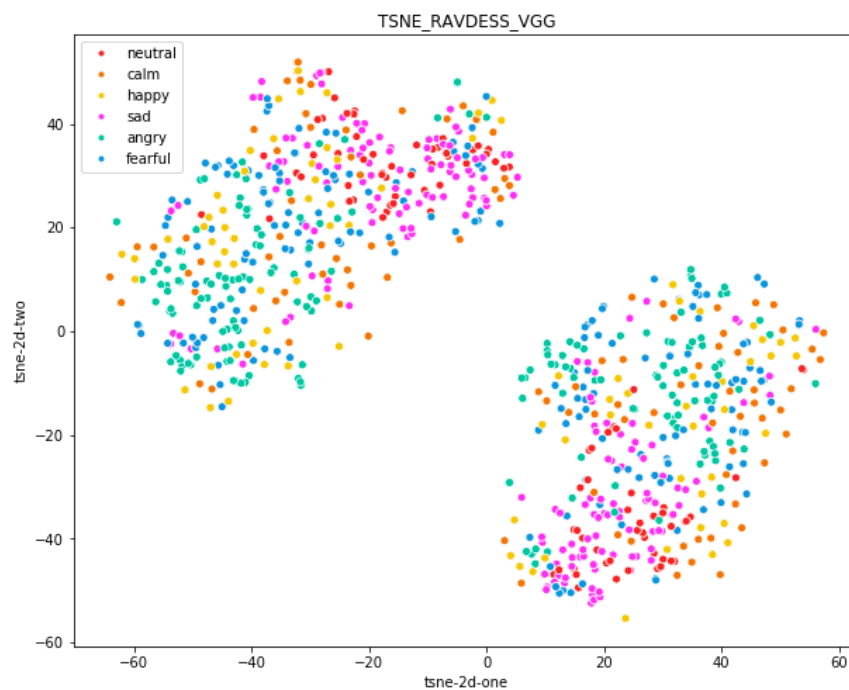

Figure 1-9: TSNE of VGGish embeddings on RAVDESS dataset (clusters are located in the left/right sides of the figure based on female/male singer)