Judgments of timbre similarity across instruments and playing techniques

Mathieu Lagrange, Vincent Lostanlen, Joakim Andén

November 2, 2020

Inter-instrument similarity

We ask whether human listeners tend to cluster audio samples by instrument. To answer this question, we derive from the consensus clustering graph $\mathcal{G}_0$ an instrument-wise similarity matrix $\mathbf{A}_\mathcal{I}$ whose rows and columns are defined on the finite set $\mathcal{I}$ of instruments. For every instrument–instrument pair (i,j), we set the value of $\mathbf{A}_\mathcal{I}$ at row i and column j equal to the number of edges in $\mathcal{G}_0$ connecting IMTs from instruments i and j, renormalized by the number of IMT samples from either instrument i or instrument j.

Let us denote the existence of an edge in $\mathcal{G}_0$ connecting two IMT samples m and n by the binary relation $m \overset{\mathcal{G}_0}{\sim} n$. The mathematical definition of the matrix $\mathbf{A}_\mathcal{I}$ is:

$$\mathbf{A}_\mathcal{I}(i,j) = \frac{\mathrm{card}\big\{(m,n) \,|\, \mathrm{Instrument}(m) = i; \mathrm{Instrument}(n) = j; m \overset{\mathcal{G}_0}{\sim} n\big\}}{\mathrm{card}\big\{n \,|\, \mathrm{Instrument}(n) \in \{i,j\}\big\}}. \quad (1)$$

Then, we run Ward's method [**?**] to cluster instruments in $\mathcal{I}$ according to the similarity matrix $\mathbf{A}_\mathcal{I}$. This method yields a permutation $\sigma_\mathcal{I} : \mathcal{I} \rightarrow \mathcal{I}$ of the rows and columns in $\mathbf{A}_\mathcal{I}$ such that the distance $|\sigma_\mathcal{I}(i) - \sigma_\mathcal{I}(j)|$ after permutation is small if and only if the similarity $\mathbf{A}_\mathcal{I}(i,j)$ is large. Figure 1 displays the rearranged similarity matrix $\widetilde{\mathbf{A}}_\mathcal{I} : (i,j) \in \mathcal{I}^2 \mapsto \mathbf{A}_\mathcal{I}(\sigma_\mathcal{I}^{-1}(i), \sigma_\mathcal{I}^{-1}(j))$ as a result of such agglomerative hierarchical clustering procedure.

Interestingly, the matrix $\widetilde{\mathbf{A}}_\mathcal{I}$ exhibits a block diagonal structure, which roughly reflects the classical taxonomy of musical instruments: four woodwinds (BbCl, Fl, ASax, Ob) are clustered together, followed by a cluster containing one woodwind (Bn) and four brass (BBTb, Hn, TTbn, and TpC), followed by a cluster of all strings (Cb, Va, Vn, Vc, Gtr, Hp). Furthermore, the largest cross-instrument similarity arises for the only pair of purely plucked instruments, namely, harp and guitar.

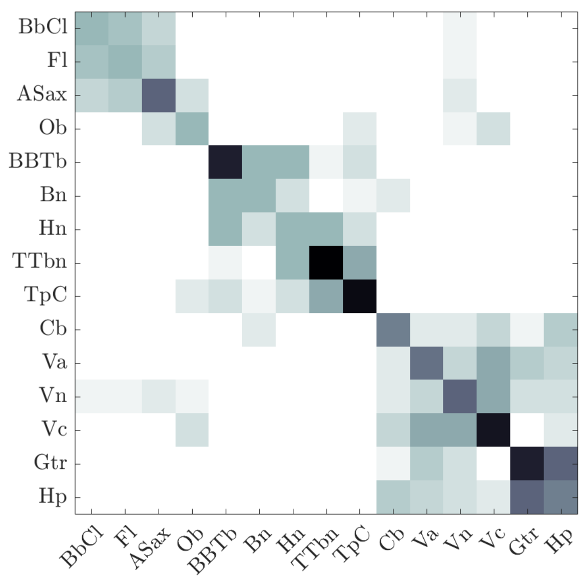

Figure 1: Matrix $\widetilde{\mathbf{A}}_{\mathcal{I}}$ of perceived timbre similarities between instruments from our dataset of $N = 78$. samples. Darker shades indicate higher frequencies of co-occurrence in the cluster graph $\mathcal{G}_0$ (see Equation 1), obtained from a consensus of $K = 31$ subjects. The rows and columns in $\widetilde{\mathbf{A}}_{\mathcal{I}}$ were re-arranged according to Ward's minimum variance method. Observe that the block diagonal structure in $\widetilde{\mathbf{A}}_{\mathcal{I}}$ reflects an organological taxonomy of musical instruments: woodwinds, brass, and strings.

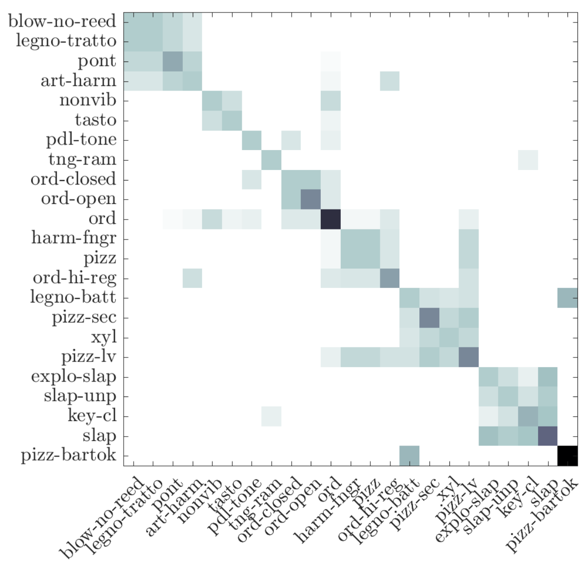

Figure 2: Matrix $\widetilde{\mathbf{A}}_{\mathcal{T}}$ of perceived timbre similarities between playing techniques from our dataset of $N = 78$ samples. Darker shades indicate higher frequencies of co-occurrence in the cluster graph $\mathcal{G}_0$ (see Equation 2), obtained from a consensus of $K = 31$ subjects. The rows and columns in $\widetilde{\mathbf{A}}_{\mathcal{T}}$ were re-arranged according to Ward's minimum variance method.

Inter-technique similarity

From the consensus clustering graph $\mathcal{G}_0$, we also derive a technique-wise similarity matrix $\mathbf{A}_{\mathcal{T}}$ whose rows and columns are defined on the finite set $\mathcal{T}$ of playing techniques. As for $\mathbf{A}_{\mathcal{I}}$, we define the similarity between any two playing techniques u and v in $\mathcal{T}$ as the following ratio of set cardinalities:

$$\mathbf{A}_{\mathcal{T}}(u, v) = \frac{\mathrm{card}\big\{(m,n) \,|\, \mathrm{Technique}(m) = u; \mathrm{Technique}(n) = v; m \overset{\mathcal{G}_0}{\sim} n\big\}}{\mathrm{card}\big\{n \,|\, \mathrm{Technique}(n) \in \{u, v\}\big\}}. \quad (2)$$

Again, we run Ward's method to cluster playing techniques in $\mathcal{T}$ according to their similarity $\mathbf{A}_{\mathcal{T}}$. Figure 2 displays the rearranged similarity matrix $\widetilde{\mathbf{A}}_{\mathcal{T}} : (u,v) \in \mathcal{T}^2 \mapsto \mathbf{A}_{\mathcal{T}}(\sigma_{\mathcal{T}}^{-1}(u), \sigma_{\mathcal{T}}^{-1}(v))$.

The block diagonal structure of the matrix $\widetilde{\mathbf{A}}_{\mathcal{T}}$ is less salient than in $\widetilde{\mathbf{A}}_{\mathcal{I}}$. Nevertheless, these clusters correspond to some basic attributes of qualitative timbre: by and large, the top-left region of the matrix contains sustained sounds whereas the bottom-right region contains percussive sounds. Interestingly, the ordinary technique (`ord`) appears at the center of the matrix, i.e., at the intersection between sustained sounds and percussive sounds. This is because the notion of "ordinariness" does not prescribe the same gesture for all instruments: e.g., an ordinary guitar sound is expected to be percussive whereas an ordinary flute sound is expected to be sustained.