# Visualizing Conversation Atmosphere in HSL Color Space

Akio Sashima
*Research Institute on Human and Societal Augmentation*
*National Institute of Advanced Industrial Science and Technology (AIST)*
Kashiwa, Chiba Japan
email: sashima-akio@aist.go.jp

Mitsuru Kawamoto
*School of Data Science and Management*
*Utsunomiya University*
Utsunomiya, Tochigi Japan
email: mkawamoto@a.utsunomiya-u.ac.jp

## I. Abstract

When people talk in groups, the atmosphere of their conversation can change significantly depending on their interactions. However, once the conversation is over, it's hard to remember whether it was lively or boring. If we could record this atmosphere for later review, we would better understand group conversations. This is especially important in places like care homes or restaurants where groups spend time together. It is valuable to know if everyone is having a good time, and this information could even become an important measure for these places.

Recently, AI technology for recognizing human speech has greatly improved. The spoken words are accurately recorded and can be reviewed later. However, this only captures what is said, not the overall atmosphere of the conversation. This means there is a gap between the data obtained from AI tools and truly understanding the full picture of group interactions.

In this demonstration, we present a novel system for real-time sensing and visualization of conversational atmosphere. The proposed system, which works on a laptop PC, continuously senses the environmental sound in the demonstration space and visualizing "atmospheric color," which is color representation of surrounding sound environment. The system also shows the history of the atmospheric colors graphically and plots the color points on a 2D map to represent atmospheric colors corresponding to the progression of the conversation. The system was implemented using Python and librosa [1], an audio processing library for Python.

## II. Technical Details

This demo system presents real-time visualizations of the sensing environment based on atmospheric colors. In this study, the atmospheric color is generated through the following process: first, the system performs environmental sound analysis to capture three sound features from human conversations: chroma pitch pattern, centroid of chroma pitch patterns, and sound loudness. Then, these features are mapped to the corresponding three color attributes, Hue, Saturation, and Lightness, and are shown as atmospheric colors in the HSL color space. The detailed process of sound feature extraction and mapping is explained as follows:

### A. Sound Feature Extraction

#### 1) Conversation Detection as Filtering Sound

To detect conversational segments, the demo system first obtains one-second audio segments from the input audio buffer, which is updated by the laptop's microphone. Then, the system applies YAMNet [2], a DNN model for sound event recognition, to each one-second audio segment and adopts the segment as a conversation part if YAMNet classifies it as "Speech."

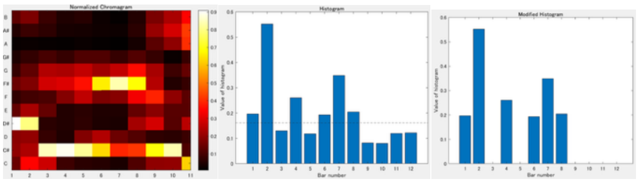


Fig. 1 Process of generating modified histogram: a) Chromagram of sound (left), b) Histogram of chroma pitch values (center), and c) Modified histogram with values below the median pitch value set to zero (right).

#### 1) Generation of Modified Histgram

Fig. 1-a shows the chromagram of a one-second conversation segment. Using this chromagram, a histogram of the frequency components is calculated, where the dashed line in Fig. 1-b represents the median value of the histogram. To emphasize the characteristics of the histogram, a modified version of the histogram is created in which values below the median are replaced by zeros.

#### 2) Normalized Centroid as Color Saturation:

The degree of saturation is assumed to be the normalized centroid of the modified histogram shown in Fig. 1-c. Specifically, the normalized centroid is calculated using the following equation.

$$\mathrm{NC} = \left(\sum_{k=1}^{N} \mathrm{f(k)} * \mathrm{hv(k)} \,/\, \sum_{k=1}^{N} \mathrm{hv(k)}\right) /\mathrm{N}$$

where NC denotes the normalized centroid of the modified histogram, N denotes the maximum number of bars, i.e., N = 12, f(k) denotes bar number, i.e., f(1) = 1, f(2) = 2,…, f(12) = 12. and hv(k) denotes the value corresponding to the bar number k.

#### 3) Chroma Pitch Pattern as Color Hue

For each bar in the modified histogram shown in Fig. 1-c, a value of one is assigned if the bar's value is greater than zero, otherwise, it remains zero. These 12 binary values are combined into a 12-bit binary number, with bar number one as the least significant bit (LSB) and bar number 12 as the most significant bit (MSB). For example, the histogram shown in Fig. 1-c corresponds to the binary value 000011101011. This binary value is converted to a scalar, which is then normalized by dividing it by the maximum 12-bit binary value: 4,095. The resulting normalized scalar is assumed to represent the degree of hue.

### 4) *Loudnes as Color Lightness*

The degree of lightness is determined by the loudness level of the one-second speech data. We create Bark-scale spectrograms for the data, and the maximum value within the spectrogram matrix is assumed to represent the degree of lightness.

### 5) *Converting Atmospheric Color*

The system assumes that the three obtained values for each one-second conversation segment represent the atmospheric color in the HSL color space. In other words, the hue, saturation, and lightness values directly correspond to the HSL components and are converted into an RGB color value. The conversion is performed using a color system library in Python.

## B. *Demo System*

Fig. 2 shows a screenshot of the demo system. The large window on the right includes the chromagram of one second of conversation data (top), the histogram of the chromagram above with the three HSL color values (second from the top), the spectrogram used to calculate the lightness value (third from the top), and the history of atmospheric colors (bottom). The upper left window is a control panel, and the lower left window shows a two-dimensional map of the atmospheric color history. These windows are updated repeatedly every one second.

## III. Relevance

This work proposes a novel direction in signal processing techniques motivated by the goal of understanding the atmosphere of a conversation. The main idea is the introduction of atmospheric color. There has been research on representing sound through colors; for example, Nagata et al. [3] attempted to identify which sounds correspond to which colors based on human synesthesia and suggested that nonverbal sound-color mappings exist latently in the general population even without synesthesia. Thus, we believe that our approach is an important step toward human-friendly visualizing the sound environment. It also opens up new possibilities for research in fields like environmental sound analysis, human-computer interaction, and social communication analysis. As this approach is still under validation, improving feature extraction and mapping is crucial, and these remain ongoing research challenges in signal processing.

## IV. Logistics

Since the demo system runs on a laptop, a desk and a power strip would be required. We would use a monitor larger than the laptop display if available.

## Acknowledgment

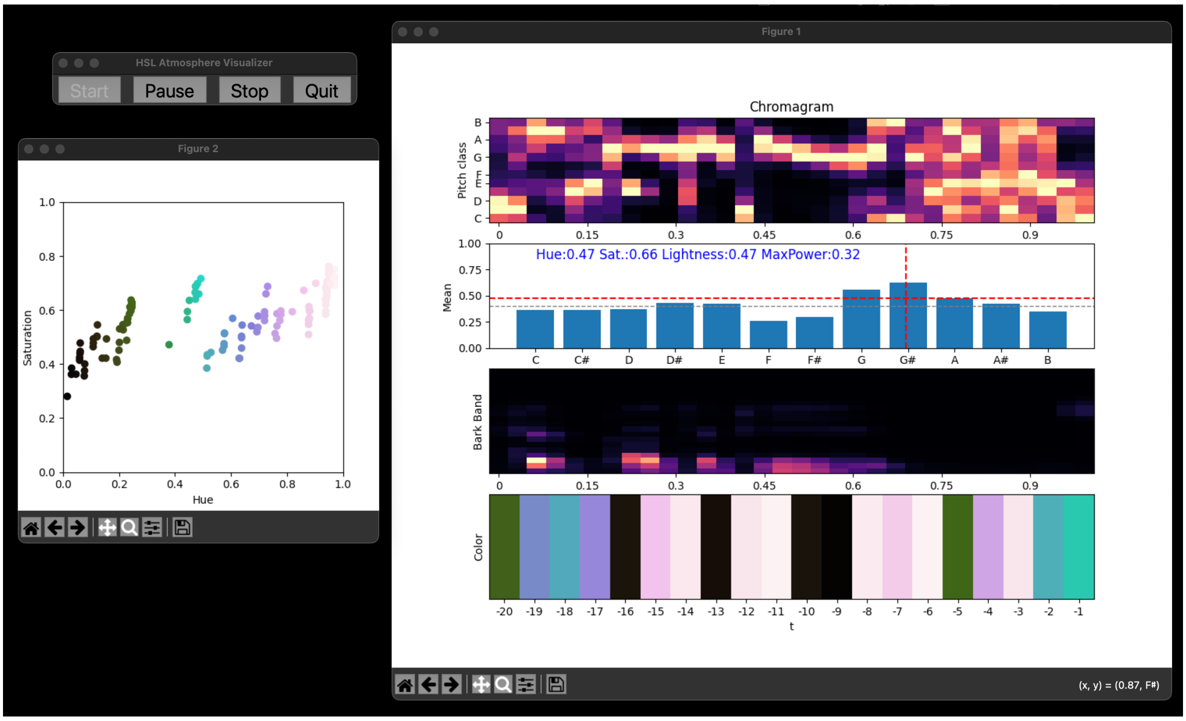
This work was partially supported by Council for Science, Technology and Innovation, “Cross-ministerial Strategic Innovation Promotion Program (SIP), Development of foundational technologies and rules for expansion of the virtual economy,” (JPJ012495) (funding agency: NEDO).

## References

[1] B. McFee, C. Raffel, D. Liang, D. P.Ellis, M. McVicar, E.Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in python,” *Proceedings of the 14th python in science conference*, Vol. 8, 2015.

[2] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, and R. C. Moore, “Audio Set: An Ontology and Human-Labeled Dataset for Audio Events,” *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2017, pp. 776‑80. *DOI.org (Crossref)*, doi:10.1109/ICASSP. 2017.7952261.

[3] N. Nagata, D. Iwai, S. Wake, and S.Inokuchi, “Non-verbal Mapping between Sound and Color—Mapping Derived from Colored Hearing Possessors and Its Applications,” *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. J86-A, No.11, pp.1219-1230, 2003 (Japanese).



Hue:0.47 Sat.:0.66 Lightness:0.47 MaxPower:0.32

Fig. 2. A screenshot of the demonstration system.