Generating music from literature using topic extraction and sentiment analysis

Jessie Salas

This article presents Tambr, a new software for translating literature into sound using multiple synthesized voices selected for the way in which their timbre relates to the meaning and sentiment of the topics conveyed in the story. It achieves this by leveraging a large lexical semantic database to implement a machine-learning-based synthesizer search engine used to select the synthesizers whose meaning best reflects the ideas of the novel. Tambr uses sentiment analysis to generate the pitches, durations, and intervals of the output melodies in a way corresponding to the sentiment of the novel—implementing algorithmic composition of literature-based music at a level of musicality not previously explored.

On timbre

Musical timbre is one of the most defining characteristics of how a piece of music sounds. Timbre refers to the character and quality of a sound, as opposed to the pitch or loudness. A saxophone and an electric guitar, for example, could play the same notes, but would still be distinguishable as different instruments; that difference is in their timbre. Because of the large number of features that define the timbre of a sound, be it from an acoustic musical instrument or a computer-generated synthesizer, academic discussion of timbre is often limited. Computational approaches to analyzing timbre are still in thier early stages; seminal studies such as the work of David Wessel and Kai Siedenburg at the Berkeley Center for New Music and Audio Technologies have explored visual ways of interpreting timbre. Other approaches, like that of Severine Samson, have explored more generative models of timbre, synthesizing musical timbre based on several parameters.

One of the most significant difficulties to approaching a discourse on timbre is the lack of consistent features and domains for describing it. A listener might describe timbre using phrases like *crunchy*, *buzzy*, *airy*, *soft*, *hard*, *rough*, and *dainty*. Such inconsistent examples of timbre make

Digital Object Identifier 10.1109/MPOT.2016.2550015
Date of publication: 10 January 2018

Tambr uses sentiment analysis to generate the pitches, durations, and intervals of the output melodies in a way corresponding to the sentiment of the novel—implementing algorithmic composition of literature-based music at a level of musicality not previously explored.

the study of it highly subjective, and an exhaustive generative mapping of timbre space is out of the scope of this project. The machine- learning algorithm used by Tambr leverages a consistent human-labeled description of synthesizer timbres, which will be discussed later in detail.

On generating music from literature

The first work describing a process of generating music from literature was published in 2014 by Hannah Davis with her system TransProse. This article describes Tambr, an advancement that builds on the approach first explored by TransProse. The central contributions of Tambr are as follows:

- Tambr uses up to 12 simultaneous voices to synthesize music, while TransProse used only piano.
- Tambr uses natural language processing algorithms to extract the topics in the novel
- Tambr varies the intensity of notes based on sentiment analysis
- Tambr follows the thematic arch and climax of the novel.
- Tambr takes musical timbre into account when selecting voices.

Tambr works toward the same goal as TransProse, taking an approach more centered on how the instruments sound. This attempts to push the current state of the art in algorithmic composition from literature by utilizing new techniques informed by standard methods used in natural language processing to get closer to the underlying narrative.

Processing the novel

The input to Tambr is a text file containing a piece of literature. The text can be any size, though larger texts tend to work better with the topic extraction algorithm used. The next step after reading the text file is to extract the relevant topics from the novel.

The program performs a term frequency–inverse document frequency (TF-IDF) transform on each paragraph of the text, which provides a weight of how relevant each word is to a given paragraph based on the number of other paragraphs in which it appears. This makes common words such as *the*, *a*, *of*, and *on* lower-weighted than more potentially defining words, which correspondingly become higher-weighted. The program represents the TF-IDF-transformed values in a document-term matrix and performs a nonnegative matrix factorization on the matrix, which can help uncover latent (hidden) relationships and topics not explicitly present in a data set.

The output of this process is a list of topics, where each topic is a set of, at most, ten terms that define the topic. For example, {*caterpillar*, *Alice*, *king*, *hatter*, *queen*, *gryphon*} is one of the topics extracted from Lewis Carrol's *The Adventures of Alice in Wonderland*. The topic extraction algorithm itself does not generate labels for the topic; it only returns a group of words that define it. If a person interprets this example topic, he/she might recognize that it is a list of central characters in the novel. Another example of a topic extracted is {*war*, *party*, *power*, *doublethink*, *society*}, which was taken from George Orwell's *1984*. We can see that this topic characterizes the defining traits of the dystopian society described in the novel.

Selecting the synthesizer

The database of synthesizers was taken from Apple Logic Pro X's (digital audio workstation) synthesizer list, and it included 650+ synthesizers, which were separated by category as per Logic's default categorizations: electronic dance music bass, percussion, plucked, rhythmic, soundscape, strings, pad, lead, classics, brass, bell, and bass. There is no large explicitly human-labeled database of how synthesizers sound in terms of their timbre, but the exploitation is here: Tambr uses the names of the synthesizers as a consistent set of labels indicating how the synthesizers might sound. The naming convention used by Apple Logic Pro X is consistently descriptive, with synthesizers named {Alien Alarm, Fog Machine, Sheets of Metal, Robot Talk, Noise Pump, Laser Brain, Dark Movements, Black Sun, Distant Air, Flying Waves, and Peaceful Meadow}.

It is now this synthesizer-selection module's job to assign the synthesizer with the label most similar in meaning to each topic previously extracted. In this novel approach, Tambr implements a search engine for synthesizers based on a large semantic database made popular by Google, as well as an open-source tool, word2vec.

Every word has a context, and words with similar contexts probably have similar meaning. Take these contexts for example:

- I am going to eat a *pie.*
- I am going to eat a *burrito.*
- I am going to eat a *banana.*
- I am going to eat a *burger.*
- I am going to eat an *X.*

These words have similar contexts, so a reader can infer that although we do not know what *X* is, it probably is a food.

Google word2vec is a two-layer artificial neural network whose output allows us to represent the entire contexts of words as a series of numerical values in a manageable vector space, making it possible to effectively process the meaning of words with computational methods. One simple way of computing how similar words are, now that we have their context-vectors (called *word-embeddings* in the natural language process literature), is to compute the cosine distance between them. Words that have cosine distances very close to zero are words that are likely synonyms of each other.

Tambr leverages this fact in its synthesizer search engine and computes the pairwise cosine distance between the words characterizing a topic of the novel and the words characterizing the timbre of a synthesizer. The computation that results in the smallest distance is the one with the best match. See Fig. 1 for a graphic depicting synthesizers paired with topics extracted from George Orwell's novel *1984*.

The sounds generated by this system are very characteristic of ambiance music, as the emphasis was placed on selecting musical timbres that match with the themes of the novel, rather than compositional rules.

Generating music

We can perform sentiment analysis using Syuzhet, an open-source library, on each sentence of the novel and get a numerical sentiment rating ranging from −1.5 to 1.5. This sentiment rating informs how Tambr generates music, given what we know from the music perception literature, in these ways:

- Quieter volume is sadder—louder volume is happier.
- Slower tempo is sadder—faster tempo is happier.
- Lower pitch is sadder—higher pitch is happier.

Tambr generates the sequence of pitches based on a linearity of dissonance proposed by Pantelis N. Vassilakis in 2007, defining dissonance in terms of roughness, illustrated in Fig. 2. The more negative the sentiment at a particular point in the novel, the more dissonant or rough the intervals in the chords will be.

The sentiment analysis process interprets the sequence of sentiment scores for each sentence as a series of signals and performs a Fourier transform with a low-pass filter to attain a smooth arch defining the dramatic arches of the novel. In Fig. 3, we see the shifts in sentiment that characterize George Orwell's novel *1984*. The loudness of the notes were defined based on this function.

Once Tambr attains the distribution of sentiment, it takes the statistical variance of the entire list. Novels with high sentimental variance mean dramatic plot shifts and more emotional density. This is accounted for by making the number of notes selected in a chord-voicing larger for novels with more emotionally varying sentiment. (The result can be found at http://tambr.ml/.)

Discussion and takeaways

As illustrated in Fig. 3, the sentiment arches generated by the sentiment analysis module correspond in a sensible way with the plot structure that a reader can verify in the text. Subjective evaluation of

FIG1 Each colored bubble corresponds to a topic extracted from Orwell's *1984*, where topics are represented as groups of related words and phrases in terms of their context within the novel. Tambr matches each topic with a synthesizer, whose name is displayed in this figure by its corresponding topic.

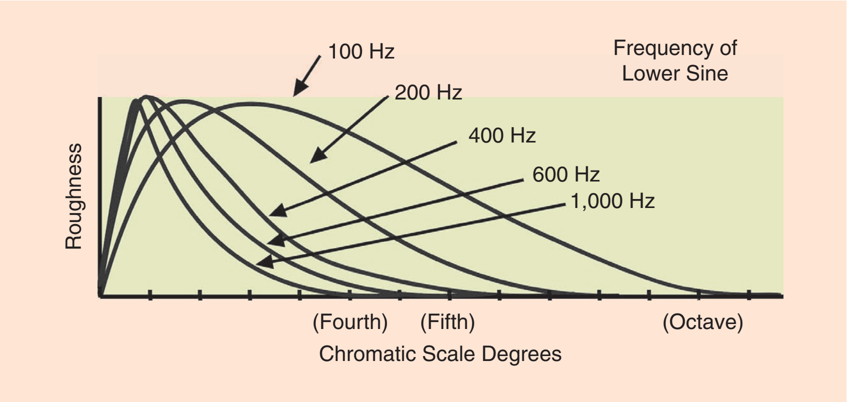

FIG2 The Vassilakis linearity of dissonance. This model gives a roughness score to each interval, which Tambr pairs with the sentiment of the text to determine the tonality of the sounds generated.

Tambr computes the pairwise cosine distance between the words characterizing a topic of the novel and the words characterizing the timbre of a synthesizer.

the tonalities produced when these sentiments are paired with the Vassilakis linearity of dissonance, and how well the resulting chord structures capture this sentiment when translated into sound, are left to the listener.

The sounds generated by this system are very characteristic of ambiance music, as the emphasis was placed on selecting musical timbres that match with the themes of the novel rather than on compositional rules. Contrastingly, TransProse placed more emphasis on the melodic content, so I chose to push the unexplored part with this first iteration of Tambr rather than re-implement the existing methods.

Further iterations of Tambr should be more intentional and less random with melody generation, along with the already intentional chord quality and timbre selection. The code for Tambr is hosted on an open-source repository accessible from the website, hopefully contributing to further projects intending to translate novels into music and to the body of knowledge of algorithmic composition as a whole.

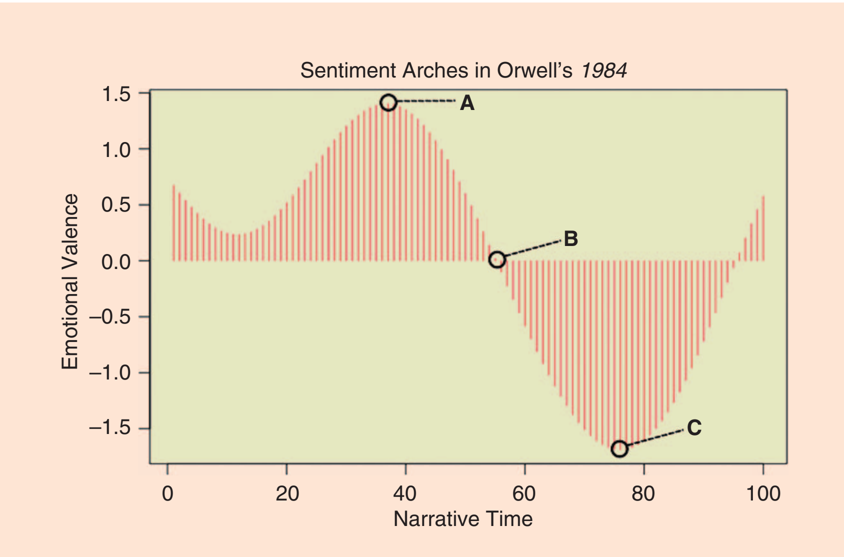

FIG3 A graph of the emotional valence (sentiment) distribution for Orwell's *1984*. The peaks and valleys represent the turning points of sentiment (and, accordingly, plot) throughout the novel. Point A (Part Two, Chapter 1) is the point of most positive sentiment before the story turns. Here, Julia secretly gives Winston a letter that reads, "I LOVE YOU." Point B (the end of Part Two, Chapter 7) is where the story reverses from generally positive sentiment to negative sentiment. This point corresponds to the scene in which Julia and Winston discuss the dangers of betraying the totalitarian political regime but profess that, regardless of what happens, the government could never remove their love for each other. Point C (Part Three, the end of Chapter 1) is the part of the story most characteristic of negative sentiment. It marks the point where Winston discovers that O'Brien, whom he had previously thought to be a member of the resistance, is actually the chief operator of The Ministry of Love (in charge of torture and law enforcement) and was posing undercover as a member of the resistance in to capture those who oppose Big Brother.

Read more about it

- K. Siedenburg, I. Fujinaga, and S. McAdams, "A comparison of approaches to Timbre descriptors in music information retrieval and music psychology," *J. New Music Research*, vol. 45, no. 1, 2016.
- H. Davis and Saif M. Mohammad, "Generating music from literature," in *Proc. EACL Workshop on Computational Linguistics for Literature*, 2014.
- P.N. Vassilakis, "Representing sound energy, phase, and interference using three-dimensional signals" presented at the 153nd meeting of the Acoustical Society of America, Salt Lake City, UT, 2007

About the author

Jessie Salas (jessie.salas@berkeley.edu) is currently a machine-learning engineer at Apple. He has a background in cognitive science and computational modeling from the University of California (UC), Berkeley. While at UC Berkeley, he cotaught the courses Data Science and the Mind as well as Introduction to Jazz Improvisation. He was involved in research at the Berkeley Institute for Data Science, Jacobs Institute for Design Innovation, and the Language and Cognition Lab. His research interests lie in how quantitative disciplines such as data science and computation can interact with art, design, and language.

P