

Wahrheit. Freiheit. Dummheit. Bosheit. (Truth. Freedom. Stupidity. Malice.)

Oded Ben-Tal

Department of Performing arts
Kingston University
o.ben-tal@kingston.ac.uk

Abstract

This piece – a set of songs for a singer and fixed electronic part – creatively explores a chain of transformations between words and music. The first link in the chain is – music inspired by words – Schumann’s song cycle *Dichterliebe* based on poems by Heinrich Heine. The next link are excerpts from textual descriptions of these songs. I ‘translated’ those back into music by using the words as prompts for text-to-music machine learning models. I treated the resulting soundfiles as sources for composing the fixed electronic part of these songs. The text for my songs is also by Heine, but different poems to those Schumann chose for his song cycle. The piece adds an imaginative contribution to the vagaries of music-text relationships by injecting the opacity of Large Language Models when text prompts are used for music generation.

1 CONTEXT

Wahrheit. Freiheit. Dummheit. Bosheit. (Truth. Freedom. Stupidity. Malice.) is a set of songs that incorporate prompt-based machine learning into a chain of transformation between words and music – an unreliable process which opens creative possibilities. The scholarship on the relationship between music and words is extensive and wide ranging (Clayton 2008). This piece creatively explores two particular corners of this wide field: setting words to music and using words to describe music. Verbal descriptions of music are used as prompts for a generative AI system resulting in very different music. This (mis)translation process is embedded within the context of setting words to music at either end: the piece is a set of songs and the descriptive texts are about Schumann’s song cycle *Dichterliebe*.

Describing music in words is a notoriously difficult problem (Pedersen, 2008. Sharma, Frank, & Zotter, 2015). Sturm et al (2014) show that the problem itself is not well defined and propose to formalise it in order to enable meaningful computational research. Tanner and Budd (1985) argue that unlike other art forms e.g. painting or literature, music is not contiguous with everyday experiences in that regard. We recount stories, talk about people, objects, and scenarios, we reference colours and shapes in conversations. But we rarely, if ever, have reason to describe, in words, sound qualities, motivic transformations, or tonal relationships outside of music. Verbal descriptions (such as programme notes) are supposed to fill this gap and aid or guide concert audiences, though it is not clear that this is always the case (Hellmuth Margulis 2010). Programme notes often contain three types of information: technical, descriptive, and contextual. (Blom, Bennett, & Stevenson 2020). They do note, though, that conjunction between these categories is common: e.g. technical details

underpinning descriptive texts. For the purposes of text-to-music systems, all three types of information should influence the generative process. However, whether the machine ‘interpretation’ of this information aligns with human listeners’ is doubtful.

2 COMPOSITION PROCESS

Generative machine learning models are notoriously opaque: the relationship between user input and system output is difficult to decipher. In effect, text-to-music models compound the inherent difficulty of describing music with the opacity of a large, complex computational model. When composing this piece I used such a model as one link in a chain of transformations back and forth between music and words. The first link in the chain is Schumann’s *Dichterliebe* – settings of poems by Heine (music ← text). I collected textual descriptions (text ← music) of that song cycle from a range of sources, and selected descriptive and evocative statements such as:

- Tears have transformed to flowers, and the vague, general longing expressed at the outset is shortly to become ecstatic embrace of “the one”
- Two dramatic crescendos push the music and the poem to a climax at the words “Herzen.” The pulsating rhythm gives a forward feeling while the dotted rhythm in the voice provides a sense of vigor.
- This disquiet continues in the closing phrases of the fourth song, [Wenn ich' in deine Augen seh',] as the piano begins to sound the dichotomy of the poet's wish and his disbelief in its fulfillment.

For the next link in the chain I turned to generative AI (mostly Stable Audio¹ version 2.5): I used the textual descriptions as prompts for music generation (music ←

¹ <https://stability.ai/stable-audio>

text). In many cases, and more so initially, I added additional wording to drag the model away from the default pop-song mode. I experimented with phrases such as ‘unusual harmonies’ ‘environmental sounds’ as well as ‘instrumental only’.

I used the resulting audiofiles which, naturally, had nothing to do with the music the prompts originally described, as sources for the fixed (‘tape’) part for the piece. I turned to the poetry of Heine myself for the text of my songs (music ← text), though very different in tone from those Schumann chose. The piece, therefore, is an example of an AI-in-the-loop co-creative process.

As Choi et al (2025) observe, communication with the model was difficult, in part because musical terminology was mostly useless. As a result, I had to alter my workflow to adapt to the limitations of the technology. In previous work with electronic medium I shaped the material iteratively guided by some initial sonic-aesthetic goals; Goals which themselves sometimes shifted during the search process. In contrast, here the starting point was procedural – using textual descriptions of music as prompts for a generative AI model. I could not predict the outputs I will be hearing. I generated over 100 short (mostly 60-90 seconds) soundfiles which I then used as ‘found’ sounds to construct the fixed electronic accompaniment to the voice. I used collage techniques, selecting and rearranging sound snippets, layering and juxtaposition but did not apply radical transformation of the source material. Surprisingly, some of the generated outputs deviated from the prevalent a=440Hz tuning. Some of these were left unchanged but in other cases I ‘corrected’ the tuning – in both cases the decision to change or not related to the vocal line, where I employ microtonal inflections. It would have been interesting to see if adding some theoretical knowledge into the generative model as suggested by Melechovsky et al (2024) would have aided the creative process or, perhaps counter-intuitively, limited some of the unexpected results.

3 PROGRAMME NOTES

The piece – five songs with a fixed electronic accompaniment – explores a chain of transformations between text and music. The first link in the chain is Schumann’s *Dichterliebe* – settings of poems by Heine (music ← text). The second link consists of textual descriptions of that song cycle, and in particular phrases that describe the music (text ← music). For the next link in the chain I turned to generative AI tools – I used the descriptions as prompts for music generation (music ← text). The resulting audiofiles, naturally, did not resemble Schumann’s music in any way, shape, or form. Nevertheless, I found enough interesting material in there to compose the fixed (‘tape’) part for the piece. Heine’s poetry is also the source text for the songs (music ← text). But rather than the lyric poetry that Schumann (and other composers) chose, I turned to his politically charged poems. These resonate powerfully with our current time across a gap of almost two centuries. Considering that Heine himself was inspired by music, among other things, when he wrote his poetry we can summarise the chain of transformation as:
Ben-Tal(AI(Writers(Schumann(Heine(...))))Heine(...)).
And if we include you, as you read these programme notes, we arrive at:

Reader(Ben-Tal(Ben-Tal(AI(Writers(Schumann(Heine(...))))Heine(...))))

4 BIOGRAPHY

Oded Ben-Tal is a composer and researcher working at the intersection of music, computing, and cognition. His compositions range from purely acoustic pieces, to interactive, live electronic pieces and multimedia work. In recent years he is particularly interested in the interaction between human and computational creativities: applying deep learning techniques to folk musics and interrogating the creative capacity of the resulting generative system within the folk tradition as well as outside it; Using AI-inspired approaches in the domain of interactive, live electronic music and joint improvisation between human performers and a semi-autonomous AI system. His work has been supported by grants from the UK’s Arts and Humanities Research Council, the Leverhulme Trust and the Volkswagen Foundation. He is an Associate Professor in the Department of Performing Arts, Kingston University, London.

5 TECHNICAL SPECIFICATIONS

The setup only requires stereo speakers to project the fixed electronic part and a microphone for the singer. The amplification of the singer should be minimal just to blend better with the electronic sounds.

Links: fixed [electronic part](#) and the [score](#) (for baritone singer. I can provide a version for mezzo as well).

AUTHOR DECLARATIONS

As outlined above, generative AI models - Stable Audio 2.5 - were used as part of the composition process. Generative AI was not used for writing this text.

REFERENCES

- Blom, D., Bennett, D., & Stevenson, I. (2020). Developing a framework for the analysis of program notes written for contemporary Classical music concerts. *Frontiers in Psychology*, 11, 376.
- Clayton, M. (Ed.). (2008). *Music, words and voice: a reader*. Manchester University Press
- Hellmuth Margulis, E. (2010). When program notes don’t help: Music descriptions and enjoyment. *Psychology of Music*, 38(3), 285-302.
- Melechovsky, J., Guo, Z., Ghosal, D., Majumder, N., Herremans, D., & Poria, S. (2024, June). Mustango: Toward controllable text-to-music generation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 8293-8316).
- Pedersen, T. H. (2008). The semantic space of sounds. *Delta*.
- Sharma, G. K., Frank, M., & Zotter, F. (2015, November). Towards understanding and verbalizing spatial sound phenomena in electronic music. In *inSonic Conference* (pp. 26-28)
- Sturm, B. L., Bardeli, R., Langlois, T., & Emiya, V. (2014, October). Formalizing the problem of music

- description. In *Int. Symposium on Music Information Retrieval (ISMIR)*.
- Tanner, M., & Budd, M. (1985). Understanding music. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 59, 215-248.
- Youjin Choi, JaeYoung Moon, JinYoung Yoo, and Jin-Hyuk Hong. 2025. Understanding the Potentials and Limitations of Prompt-based Music Generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 766, 1–15.