The Listening Machine: Generating Complex Musical Structure from Social Network Communications

Daniel Jones[1] **and Peter Gregson**[2]

Abstract. Social networks such as Twitter are a rich source of structured, expressive data. The Listening Machine is a generative music system which incorporates real-time conversations sourced from Twitter to produce a piece of data-driven music of indeterminate length, parametrised by the content, sentiment and conversational prosody of 500 online participants. This paper describes the background and composition of The Listening Machine, motivating its position as a hybrid system spanning sonification, process composition and live algorithm.

1 Background

In the early part of the 21st century, one of the dominant modes of conversation is over networked media, via text written in primarily public forums. This offers vast affordances to the analysis and understanding of discursive habits. Linguistic acts can be automatically captured and analysed for their style, sentiment and semantics, statistically associated with the speaker's demographic and geographical location [7].

These affordances have been taken up with gusto within the world of digital sociology, from general conversational analysis [10] to tracking public sentiment against stockmarket prices [4]. Less common is its uptake as a way to consider generating structure for artistic works. Though some work has been done in the area, based upon social networks specifically [6, 9] and text more generally [8, 1, 14], there has yet to be a genre-defining musical work based on the dynamics of a social network.

The authors' previous research [5, 11] focuses on incorporating real-time data sources as a way to generate and organise musical material, combining specialisms across algorithmic, data-driven composition and classical orchestration. This proposal came as an attempt to produce a thoroughgoing synthesis of the two fields, using online conversations as its structuring element.

The proposal as first conceived was to create a piece of music that represents the online behaviour of a comprehensive cross-section of the UK's online communication: effectively, a sonic portrait of internet conversations. The artistic intent was to create something which reflected the knotted network of machinic structures which comprise this new communications medium, interwoven with the human conversational "melodies" which take place upon it. The resultant work, *The Listening Machine*, was realised between January and May 2012, and publically operational from May 2012 until January 2013.

1.1 Overview

This paper describes the motivation and workings behind The Listening Machine, a digital piece of music of indefinite length which uses real-time Twitter communication as as its source of musical structure.

The objective behind the work was to create a resultant composition which:

- incorporates the conversational dynamics of a reasonable demographic cross-section of UK-based Twitter users
- expresses the dynamics, tone and rhythm of the collective conversations
- is generated in real-time, heard live over the internet
- uses orchestral musical motifs with predominantly Western tonality, scored and recorded as part of the development of the piece
- sounds as if it were composed by a human composer

The observant reader may notice an apparent paradox, or at the very least a tension: sonification and classical composition are, at first glance, orthogonal to one another. To write a piece of music means taking top-down control over its structure, phrasing and dynamics, incorporating movements and themes that are elaborated towards an ultimate teleology – traditionally, a climactic resolution of the themes [18].

Here, the overall structure is determined by an external data source. How can we still make the piece sound composed, and recognizable within the framework of Western composition? Drawing on the Minimalist compositional tradition [16], we instead introduce a macro-level structure which is progressive and based on the successive layering of individual instrument lines. It is the precise interaction between these lines which are intended to sound as if they were arranged by a human.

To give depth and engagement, mapping single speech acts to linear chains of sounds is thus insufficient. The piece requires multiple interacting layers of compositional structure which vary over time: tone and mode, multiple concurrent instrumentation and complex polyphony, responses to syntactic, semantic and rhythmic properties of speech... It also

[1] Goldsmiths, University of London, UK, email: d.jones@gold.ac.uk
[2] peter@petergregson.co.uk

requires a sufficiently robust and flexible framework to be able to perform real-time linguistic analysis, communicating with the audio workstation containing the musical elements.

Section 2 describes the toolkit used to perform this textual analysis. Subsequent sections describe the parts of the system which address each of the above compositional requirements, before progressing to discussion of the broadcast and visualisation frameworks used to transmit the music to listeners.

1.2 Systems Outline

The overall architecture of The Listening Machine is shown in Figure 1.

- The **listener** (§2) monitors a sample of online subjects, detecting new posts in real time.
- The **phonomat** (§3) analyses each text utterance along multiple axes: topic classification, sentiment, prosody, keyword triggers. It also tracks gross classifications such as sentiment and topic across all subjects, and the mean rate of conversation.
- The **conductor** (§3, §4) links textual analysis to musical elements, triggering the appropriate sound fragments based on current state.
- The **encoder** (§5) takes the output of the audio framework and encodes it for online broadcast, streamed to the internet

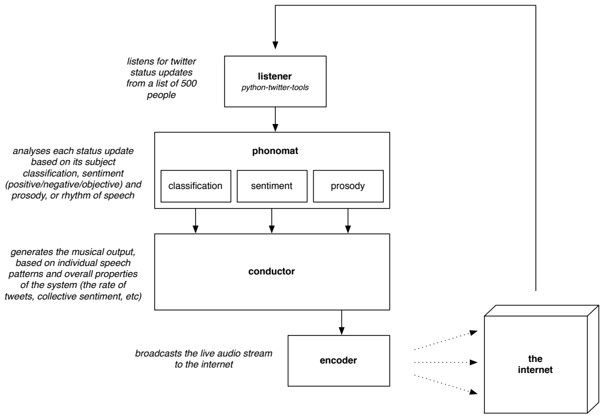

listens for twitter status updates from a list of 500 people
listener
python-twitter-tools
analyses each status update based on its subject classification, sentiment (positive/negative/objective) and prosody, or rhythm of speech
phonomat
classification
sentiment
prosody
generates the musical output, based on individual speech patterns and overall properties of the system (the rate of tweets, collective sentiment, etc)
conductor
broadcasts the live audio stream to the internet
encoder
the internet

Figure 1. System-level structure diagram

2 Data Sampling and Analysis

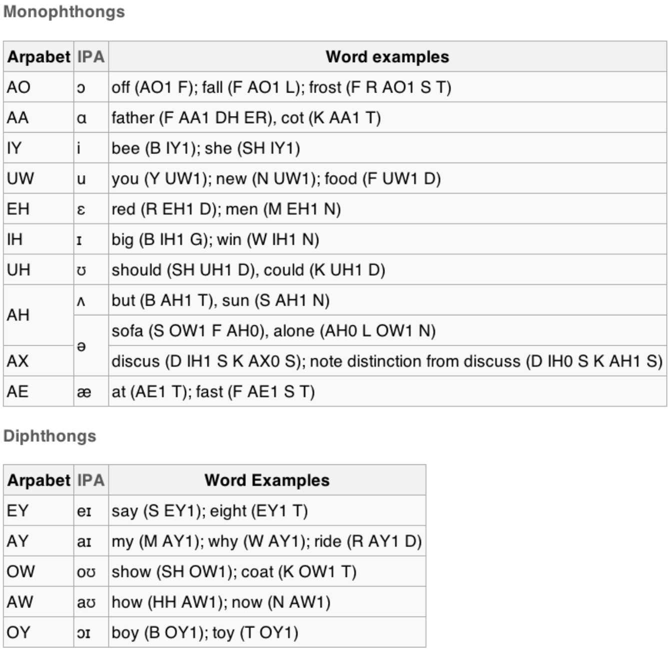
To obtain a demographically-neutral cross-section of the UK's online population, stratified cluster sampling was used. With a total target total of 500 users, 50 users each were selected from 8 different demographic categories: sports, politics, science, technology, entertainment, health, business, and arts. A further 100 were selected randomly, by sampling from searches over all UK-based tweets.

The online activity of these users is then monitored in real-time using the *Python Twitter Tools* framework, with linguistic analysis performed by the *Natural Language Toolkit* (NLTK) [3]. Key properties are extracted using NLTK, as described in the corresponding sections below.

3 Compositional Elements

3.1 Prosody

The most crucial sonification within The Listening Machine is translating a sentence's prosody – that is, its rhythm and melodic flow – into diatonic pitches. Prosody is used in many prominent examples of text-to-music or music-to-text mappings [14, 12].



Monophthongs

Arpabet	IPA	Word examples
AO	ɔ	off (AO1 F); fall (F AO1 L); frost (F R AO1 S T)
AA	ɑ	father (F AA1 DH ER), cot (K AA1 T)
IY	i	bee (B IY1); she (SH IY1)
UW	u	you (Y UW1); new (N UW1); food (F UW1 D)
EH	ɛ	red (R EH1 D); men (M EH1 N)
IH	ɪ	big (B IH1 G); win (W IH1 N)
UH	ʊ	should (SH UH1 D), could (K UH1 D)
AH	ʌ	but (B AH1 T), sun (S AH1 N)
	ə	sofa (S OW1 F AH0), alone (AH0 L OW1 N)
AX		discus (D IH1 S K AX0 S); note distinction from discuss (D IH0 S K AH1 S)
AE	æ	at (AE1 T); fast (F AE1 S T)

Diphthongs

Arpabet	IPA	Word Examples
EY	eɪ	say (S EY1); eight (EY1 T)
AY	aɪ	my (M AY1); why (W AY1); ride (R AY1 D)
OW	oʊ	show (SH OW1); coat (K OW1 T)
AW	aʊ	how (HH AW1); now (N AW1)
OY	ɔɪ	boy (B OY1); toy (T OY1)

Figure 2. Phoneme mappings in Arpabet (source: Wikipedia)

This sonification is a multi-stage process which maps each vowel phoneme to a series of different pitch mappings, targeted at different instrumentation. These mappings are designed to correlate with the frequency orderings used when speaking each vowel.

Firstly, a word is broken down into its pronunciation using the CMU Pronunciation Dictionary. This outputs a mapping in the Arpabet phonetic transcription code (Figure 2).

This transcription is then filtered for its vowel sounds, which make up the dominant prosodic perception of a word [15]. This leaves two classes of vowel sound: monophthongs (single-voiced) and diphthongs (double-voiced), as listed in Figure 2.

Each of these classes of vowel, when spoken, incorporates multiple *formant frequencies*: different relative frequency bands which give rise to the vowel's characteristic timbre. These formant frequencies – described as F1, F2 and F3, corresponding to the low, mid and high frequency bands – are

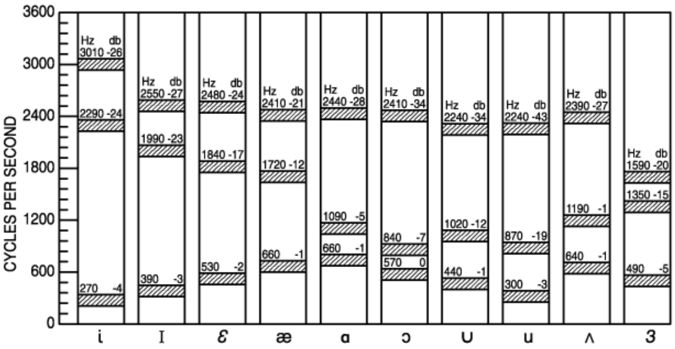


Figure 3. Formant frequencies for three components: F1, F2 and F3

used to generate orderings of pitches within a given scale. Figure 3 provides an overview of these orderings.

For example, let us assume we are in the C major scale. For the instrument playing the F1 frequencies, a vowel of *i* will generate the lowest note in the scale; that is, degree 0 (= C4). For the instrument playing F3, the same vowel will play the *highest* note in the scale; as we are categorising 10 different vowels, this is degree 9 in the scale (= D5),

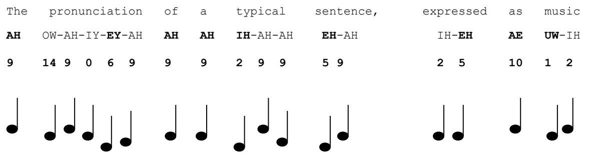

The pronunciation of a typical sentence, expressed as music
AH OW-AH-IY-EY-AH AH AH IH-AH-AH EH-AH IH-EH AE UW-IH
9 14 9 0 6 9 9 9 2 9 9 5 9 2 5 10 1 2

Figure 4. Translating a sentence to pitches of a diatonic scale

3.2 Sentiment

The second dominant property of our linguistic analysis is the collective sentiment: is the general mood positive or negative?

Various different approaches have been used to classify sentiment in Twitter data [2]. We discern sentiment using the NLTK WordNet, alongside the *SentiWordNet* database. This classifies different sentiment-expressing words into three poles: positive, negative and neutral. A significant degree of noise is inherent in such a system, due to ambiguous symbols such as those used ironically. Moreover, general population-level trends tend to be relatively minor, and only substantial during extremely major events with far-reaching emotional consequences.

To combat these tendencies, the global sentiment is normalised between $[0, 1]$, and unified across this scale by using an empirical cumulative frequency distribution. This uses the sampled noisiness of the real sentiment distribution to ensure that measured values always lie, as far as possible, uniformly across $[0, 1]$.

These positive/negative/neutral poles are then used to determine the global tonal mode used by the piece. Two modulating major modes are used for positive sentiment; minor modes for negative; and modal (phrygian, lydian) scales for neutral sentiment.

3.3 Tempo

A critical property of online conversations is their change in rate across the day. As the subjects are geographically located within one specific region, a strong daily cycle can be seen. The first utterances occur around 5am as the early risers wake up, peaking around the 9am and 5pm rush hours. As night falls, the tempo gradually recedes, with conversation becoming extremely sparse after 2am. As a highly salient quality, it is vital that this change of pace is recognizable in its musical portrayal.

Given the pre-recorded motifs that the piece is founded upon, it is difficult to use a varying tempo. To avoid altering pitch as the tempo is decreased, it is necessary to use a continuous-pitch time-stretching algorithm, which introduces substantial CPU load and audible artefacts. We thus made the decision to take another approach: fixing the tempo of the piece at 120 beats per minute, and adjusting the density of musical events to reflect the conversation rate. As the frequency of speech events drops below a certain threshold, the rate of musical events is halved, triggering subsequent patterns at 60BPM. This also serves to lengthen the expression of each sentence, leaving less idle periods when fewer conversational events are happening. Likewise, at busy periods, motifs take place in double-time, at 240BPM.

A consequence of this continuous pulse is that the piece develops a metronomic, clockwork-like rhythm, which resonates agreeably with the notion of the eponymous "Machine". The listener can almost imagine a set of gears in motion, ticking away to produce this automatic music. Specifying a tempo in multiples of 60BPM links the clock directly to the duration of a second, grounding the pace of the piece with the rate at which we, as humans, track the passing of time throughout the day.

3.4 Semantics

The stated motivation of The Listening Machine is to portray the conversations of our 500 subjects as faithfully as possible. However, there is an immediate rift introduced by the disparity between linguistic and musical expression: instrumental music has no semantics, as such. There is no way to unambiguously express the concept of "hairdryer" via a pattern of musical pitches. Directional and emotional analogues can arguably be expressed (see Section 3.2); however, this leaves most noun, verb and adjective forms omitted from the composition.

The Listening Machine incorporates two distinct approaches to remedy the problem of semantics, both based on identifying discrete subject matters: topic classification, which analyses semantic content on the population level; and keyword triggers, which looks for mentions of domain-specific terms.

3.4.1 Topic Classification

In the same vein as sentiment analysis, it is frequently useful to identify popular topics on the level of the whole population; "trending topics", in Twitter parlance. The London

Olympics, for example, led to intense, widespread coverage of sporting topics. Likewise, a general election might lead to politics becoming in the public eye.

To address this need, the linguistic analysis layer of The Listening Machine incorporates a set of topic classification algorithms, seeded with a large database of pre-classified example material. The selected material is taken from the BBC News website, which is already categorised into a set of useful categories: sports, politics, science, technology, entertainment, health, business, and arts. Using an automated system to download archives from the BBC News website based on each of these categories, and then removing any markup to produce a bank of pure text, we produced a set of UK-specific classification corpora. This is then used with a tf-idf [17] classifier to rank the weighting of each topic within a given sentence.

For example, take the statement "Financial Times headlines today: Markets showing no sign of recovery; spectators fear the worst". The terms "financial" and "markets" occur frequently within business but infrequently within general English, so this statement would strongly rank under that category. "Recovery" might flag it up as low-ranking within health, and "spectators" within sports.

Each statement can thus be normalised to rank in each category between [0..1]. Using a leaky integrator or sliding window, we can then determine the dominant conversational topics within the current time period.

A change in classification is used to activate one of a series of topic "themes": modular compositions comprised of their own multi-instrumental arrangements, with each track played according to simple chance processes. These through-composed parts herald a change of topic with its own recognizable set of melodic structures.

3.4.2 Keyword Triggers

Semantics are dealt with in a second, more literal manner: by including occasional real-world field recordings within the composition, portraying events and objects that are likely to occur within an everyday conversation. These are grouped by topic and triggered by keywords that are automatically associated with each topic. For example, a mention of the work "playground" might trigger a recording of children playing outside.

The field recordings were recorded by the artists or contributed by third parties, and are generally between 20s and 2m long, with an inward and outward fade of several seconds.

Like the topic themes, these give the listener a sense of discrete episodes within the musical material, punctuating the gradual progression of the layered prosodic parts.

4 Integrating Recorded Motifs

One of the primary intentions and challenges of this piece was to craft it from fragments recorded throughout its production with orchestral musicians, from the Britten Sinfonia chamber ensemble. Though a laborious process, this was a fundamental choice for multiple reasons – the primary being to re-humanise a portrait which is, by definition, quite dry in its process. The conversations that make up the source data are made up of feelings and emotions, which are best represented by bringing in musicians to provide a unique set of expressive performances. Through it would have been possible to use off-the-shelf sample libraries, these would have omitted the richness and nuance of a human player when (say) a series of legato notes are played on a stringed instrument. We hoped to juxtapose the steady machine-like pulse of the piece with a series of flowing recordings which give the sense of voice-like words and sentences.

We thus engaged in a series of recording sessions in which orchestral musicians played complete musical "words" corresponding to the prosody pronunciations determined previously. Multi-syllable words were analysed and scored, based on the most common word pronunciations using Arpabet (see Figure 5). These were then recorded and broken up into single-word fragments, giving a total of over 20,000 multi-syllable word "recordings".

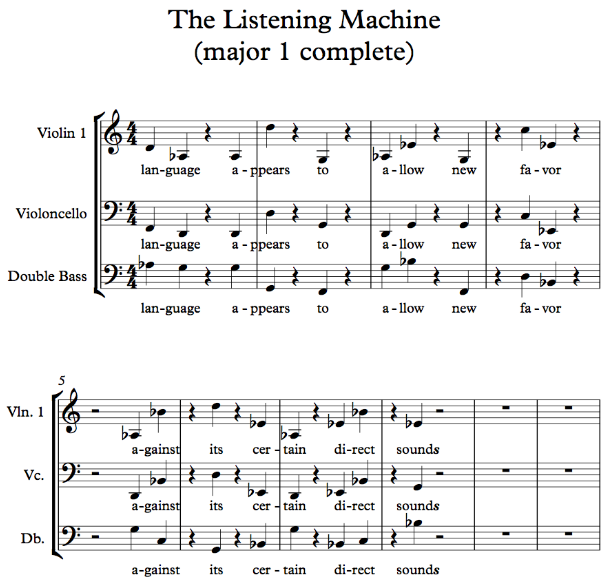

The Listening Machine
(major 1 complete)
Violin 1
Violoncello
Double Bass
lan-guage a - ppears to a - llow new fa - vor
lan-guage a - ppears to a - llow new fa - vor
lan-guage a - ppears to a - llow new fa - vor
5
Vln. 1
Vc.
Db.
a-gainst its cer - tain di-rect sounds
a-gainst its cer - tain di-rect sounds
a-gainst its cer - tain di-rect sounds

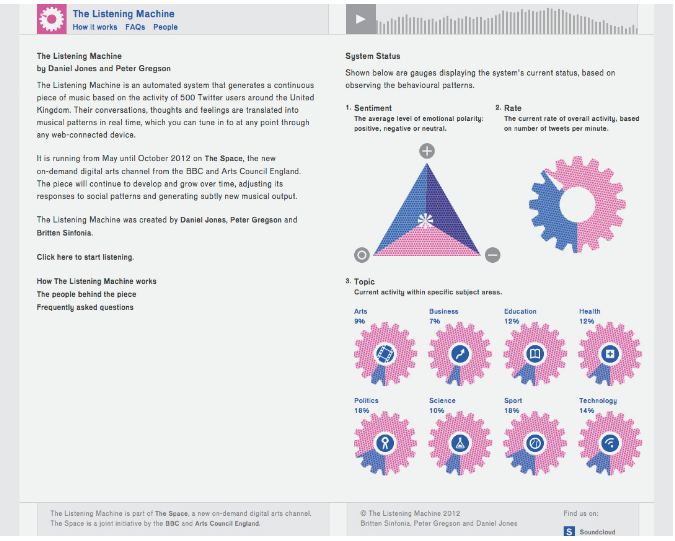
Figure 5. Excerpt from notated score for violin, cello and double bass

One of the fundamental elements of the conductor process is a thread which triggers these multi-syllable words-fragments, using a scheduling mechanism to time them correctly according to the sentence. A sentence is thus portrayed with its musical equivalent, with natural legato giving an expressive and free-flowing musical output.

5 Broadcasting and Visualisation

This piece was streamed live to a total audience of tens of thousands over a 9-month broadcast period. To do so, a streaming architecture was used which distributed a single MPEG-2 Layer 3 stream to a content delivery network hosted by the British Broadcasting Corporation.

The resultant stream was heard by viewers to the project's website, using the jPlayer HTML5/JavaScript online media streaming interface. Alongside the audio stream, the system's current status was displayed in real-time using standards-compliant HTML5 canvas rendering (Figure 6).



The Listening Machine

How it works FAQs People

The Listening Machine
by Daniel Jones and Peter Gregson

The Listening Machine is an automated system that generates a continuous piece of music based on the activity of 500 Twitter users around the United Kingdom. Their conversations, thoughts and feelings are translated into musical patterns in real time, which you can tune in to at any point through any web-connected device.

It is running from May until October 2012 on **The Space**, the new on-demand digital arts channel from the BBC and Arts Council England. The piece will continue to develop and grow over time, adjusting its responses to social patterns and generating subtly new musical output.

The Listening Machine was created by **Daniel Jones, Peter Gregson** and **Britten Sinfonia**.

Click here to start listening.

How The Listening Machine works
The people behind the piece
Frequently asked questions

System Status

Shown below are gauges displaying the system's current status, based on observing the behavioural patterns.

1. Sentiment
The average level of emotional polarity: positive, negative or neutral.

2. Rate
The current rate of overall activity, based on number of tweets per minute.

3. Topic
Current activity within specific subject areas.

The Listening Machine is part of **The Space**, a new on-demand digital arts channel.
The Space is a joint initiative by the **BBC** and **Arts Council England**.

© The Listening Machine 2012
Britten Sinfonia, Peter Gregson and Daniel Jones

Figure 6. Screenshot of The Listening Machine website, June 2012

6 Conclusion

The Listening Machine is an unashamed hybrid of multiple fields, fitting neither into pure sonification, process composition, or live algorithm. Yet it draws on each of these: it mirrors the dynamics of an external data source; it generates tonal patterns with multi-layered rhythms, extending the language of 20th Century process music; and it operates semi-autonomously, constraining its behaviour based on its current output and the web of rule systems which make up its infrastructure.

We believe that this kind of hybrid approach might help to close the gap between the intellectually-driven world of computer music and the classical music community.

Through many iterations of development, composition and careful listening, the resultant composition of The Listening Machine was remarkably varied and rich. Listeners stated that it rewarded sustained and repeated listening, commenting on the distinctive changes to timbre and structure that occur throughout the day. These anecdotes are supported by listener analytics: 11% of visitors listened for more than 3 minutes, and 4.2% for more than 10 minutes. Over 1% tuned in for upwards of half an hour. Compared to the extremely fast dropoff rate of typical web dwell times [13], we felt this was a positive indicator of sustained engagement.

The Listening Machine reached a global audience, reported in Wired, the Huffington Post, and the Wall Street Journal; Spanish daily broadsheet *El Pais* described it as “an endless and beautiful piece of contemporary music”. Part of the reward was seeing a generative, data-driven composition capturing the attention of an audience that would not normally be exposed to algorithmic music.

6.1 Future Work

Work is currently being done on translating the framework developed for The Listening Machine to a live environment, in which orchestral musicians can perform text-generated scores materialising before them in real-time.

There are a number of fruitful potential avenues which we could not explore due to ever-present time constraints. One is to incorporate a more free-flowing, non-metrical model of prosodic flow, which would allow the overall timeline of the piece to break out of the relatively strict quantization currently in place.

An alternative direction would be to introduce more complex notions of tonality and harmonic relations, such as those proposed by Tymoczko [19] or Woolhouse [20]. The current approach is functional but naïve, and may benefit from refined tonal continuity between sections, allowing for more nuanced and multidimensional expression of sentiment.

6.2 Excerpts

Excerpts from The Listening Machine can be heard at:
`soundcloud.com/ideoforms/sets/the-listening-machine-excerpts`

ACKNOWLEDGEMENTS

We would like to thank the BBC and Arts Council England for providing the funding which realised The Listening Machine, and to Britten Sinfonia as the orchestral partner whose outstanding musicianship brought The Listening Machine to life. Thanks also to the anonymous reviewers whose thoughtful criticisms have enriched this paper.

REFERENCES

[1] F. Alt, A.S. Shirazi, S. Legien, A. Schmidt, and J. Mennenöh, ‘Creating Meaningful Melodies from Text Messages’, in *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, pp. 63–68, (2010).

[2] Albert Bifet and Eibe Frank, ‘Sentiment knowledge discovery in Twitter streaming data’, in *Proceedings of the 13th International Conference on Discovery Science*, DS'10, pp. 1–15, Berlin, Heidelberg, (2010). Springer-Verlag.

[3] Steven Bird, Edward Loper, and Ewan Klein, ‘Natural language processing with Python’. O'Reilly Media Inc, (2009).

[4] Johan Bollen, Huina Mao, and Xiaojun Zeng, ‘Twitter mood predicts the stock market’, *Journal of Computational Science*, **2**(1), 1–8, (2011).

[5] James Bulley and Daniel Jones, ‘Variable 4: A Dynamical Composition for Weather Systems’, in *Proceedings of the International Computer Music Conference*, Huddersfield, UK, (2011).

[6] Luke Dahl, Jorge Herrera, and Carr Wilkerson, ‘Tweetdreams: Making music with the audience and the world using real-time Twitter data’, in *International Conference on New Interfaces For Musical Expression*, Oslo, Norway, (05/2011 2011).

[7] Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing, ‘A latent variable model for geographic lexical variation’, in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1277–1287. Association for Computational Linguistics, (2010).

[8] Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, and Shigeki Sagayama, ‘Automatic song composition from the lyrics exploiting prosody of the Japanese language’, in *Proc. 7th Sound and Music Computing Conference (SMC)*, pp. 299–302, (2010).

[9] Thomas Hermann, Anselm V. Nehls, Florian Eitel, Tarik Barri, and Marcus Gammel, ‘Tweetscapes - real-time sonification of Twitter data streams for radio broadcasting’, pp. 113–120. The International Community for Auditory Display (ICAD), (2012).

[10] Courtenay Honeycutt and Susan C Herring, 'Beyond microblogging: Conversation and collaboration via Twitter', in *42nd Hawaii International Conference on System Sciences*, pp. 1–10. IEEE, (2009).

[11] Daniel Jones, 'Atomswarm: A framework for swarm improvisation', in *Applications of Evolutionary Computing*, 423–432, Springer, (2008).

[12] David Evan Jones, 'Speech extrapolated', *Perspectives of New Music*, **28**(1), 112–142, (1990).

[13] Chao Liu, Ryen W White, and Susan Dumais, 'Understanding web browsing behaviors through weibull analysis of dwell time', in *Proceedings of the 33rd SIGIR Conference on Research and Development in Information Retrieval*, pp. 379–386. ACM, (2010).

[14] Alex McLean and Geraint Wiggins, 'Words, Movement and Timbre.', in *Proceedings of New Interfaces for Musical Expression 2009*, pp. 276–279, (2009).

[15] Sieb Nooteboom, 'The prosody of speech: Melody and rhythm', in *The Handbook of Phonetic Sciences, Nr. 5 in Blackwell Handbooks in Linguistics*, pp. 640–673, (1997).

[16] Steve Reich, 'Music as a Gradual Process', in *Audio Culture*, 304–306, Continuum, New York, NY, (2004).

[17] Gerard Salton and Christopher Buckley, 'Term-weighting approaches in automatic text retrieval', *Information processing & management*, **24**(5), 513–523, (1988).

[18] Arnold Schoenberg, Gerald Strang, and Leonard Stein, *Fundamentals of musical composition*, Faber & Faber, 1967.

[19] Dmitri Tymoczko, 'The geometry of musical chords', *Science*, **313**(5783), 72–74, (2006).

[20] Matthew Woolhouse, 'Modelling tonal attraction between adjacent musical elements', *Journal of New Music Research*, **38**(4), 357–379, (2009).