World Cloud: A Prototype Data Choralification of Text Documents

Alexandra L. Uitdenbogerd

RMIT University

Address: School of Science, RMIT University, GPO Box 2476, Melbourne VIC 3001 Australia.

Email: sandrau@rmit.edu.au

World Cloud: A Prototype Data Choralification of Text Documents

This work extends the concept of algorithmic composition to textual data and demonstrates a novel method of text setting. We attempted to represent the key words of a document set through “data choralification”, being a polyphonic singing equivalent of a word cloud. Word frequencies were determined for all words in a set of documents on the topic of sexual assault, which was the theme of a six-movement choral work, of which the choralification was one movement. Stop words were removed, then highest frequency words had their frequencies converted to audio frequencies in the normal choir vocal range. Words were allocated to vocal parts based on equal non-overlapping pitch ranges and sung in parallel. The composition was successfully performed by choristers. The impact of composition decisions is discussed, as well as how an automated choralification of data using generated synthetic voices challenges the current definition of sonification. Further work related to word selection and mood representation is proposed.

Keywords: data sonification, algorithmic composition, choral composition

1. **Introduction**

A *word cloud* is an image that contains key words of a document or set of documents, where the font size indicates the importance of the word, as determined by word frequency and the filtering of function words, such as “the”. Word clouds are related to *tag clouds,* which also use frequency to determine the font size of tags applied to items such as blog posts, providing a sense of the content of a blog, and allowing retrieval based on the tags. Word clouds have also been used for qualitative analysis of textual data (For example, Chowdhury, Koya & Philipson, 2016), and for teaching (Ramsden & Bate, 2008).

This exploratory study considers the possibility of implementing an audio version of the Word Cloud, as a form of algorithmic composition based on word frequencies of a collection of documents. Being a word cloud necessitates the use of words, in this case sung, and the representation of their frequency. An algorithmic composition for choir inspired by the concept of word clouds was generated, using a set of documents on the theme of sexual assault. The piece was one movement in a six-movement choral work on the theme of sexual assault. Its role in the work was to provide an uncomfortable text corpus summary, attempting to make sense of overwhelming information.

This work is novel in several ways. First, instead of setting a text such as a poem, an automatic summary of source text has become the text of the resulting composition, in a manner inspired by word clouds. Second, the algorithmic composition produced a choral work, which can be performed by choristers, or be generated by synthetic voices. Third, the generation of synthetically sung output from text data input challenges existing accepted definitions of data sonification.

We summarise previous work on word clouds; and discuss the relevant audio perception, information retrieval, and algorithmic composition literature in Section 2. We report in Sections 3-5 on our prototype data choralification that attempts to represent the relative prevalence of the most frequent non-function words occurring in sets of documents, and to express the underlying emotion of the documents.

2. Background

This work extends the concepts of data sonification and algorithmic composition, applying them to the text domain to produce sung word clouds. The word cloud visualisation is intended to demonstrate the relative prevalence of non-function words that have already been ranked as of sufficient importance to be included in the visualisation, and to thereby represent the main themes of the underlying source text. We also consider various extensions to the present work, such as the extraction of other aspects of text, such as style and genre. In this section we review each of these areas, as well as relevant audio perception research.

2.1 Word Clouds

Feinberg (2010) discusses the evolution of the idea of the word cloud, for which he developed the Wordle software to produce aesthetically pleasing word clouds, using a range of fonts, colours, and angles. It was inspired by tag clouds used on various websites, such as Flickr, and a graphic by Matt Jones[1]. Wordle has since been widely used, mostly for fun and decoration, but also for analysis, communication of a point, and education (Viégas, Wattenberg, and Feinberg; 2009).

Hearst and Rosner (2008) studied the application of tag clouds by interviewing information visualisers, and analysing web pages that discussed tag clouds. Tag clouds were generally seen as fun, dynamic, and linked to social interests and activity. They were not considered to be data analysis tools. However, this appears to have changed in recent years. McNaught and Lam (2010) demonstrated that word clouds (from Wordle[2]) were useful as a preliminary analysis tool, as well as a validation tool for comparing the results of multiple sources of data. Since word frequency analysis via word clouds became part of the toolkit for users of the popular qualitative data analysis software package NVivo, word clouds have appeared in articles to demonstrate themes occurring in text (For example, McMillan, Pugh, Hamid, Salinsky, Pugh,

[1] https://magicalnihilism.com/2004/07/04/my-delicious-tags-july-2004/

[2] http://www.wordle.net/

Noël, & LaFrance, 2014; Rowell, Evans-Reeves, & Gilmore, 2014; Chowdhury, Koya & Philipson, 2016).

Word clouds are not the only means of visualising the content of a body of text. Hearst (1999) has described various alternative techniques for visualising textual information and the relationships between documents, particularly as it relates to information seeking, with most techniques providing dynamic interfaces, such as zoom capability. However, word clouds are less about relationships between documents, and more about giving a quick impression about content. Despite this, there have been attempts to extend the original use of tag and word clouds to express further dimensions of data. For example, Cui, Wu, Liu, Wei, Zhou, and Qu (2010) created a system for dynamically generating word clouds along a time-line. This made it possible to see the evolution of concepts in a text data source.

Several attempts have been made to allow multiple documents to be compared via a word cloud display. For example, Lohmann, Heimerl, Bopp, Burch and Ertl (2015) developed a word cloud layout that allows a small set of documents to be compared, by having the intersection of frequent words in the centre, with document-unique words in the outer circle and others in-between. In other work by the same lab, natural language processing techniques were applied to expand the linguistic information being provided in word clouds, such as indicating the most prevalent part of speech by the colour of the word, as well as to give more nuanced control over word cloud content (Heimerl, F., Lohmann, S., Lange, S., & Ertl, 2014).

Despite the inspiration behind Wordle being the aesthetics of Matt Jones' original graphic, research into the aesthetic side is limited. A noteable exception is the work of Marszałkowski, Rusiecki, Drozdowski, and Narożny (2014), who implemented some existing typography aesthetic guidelines, including *tonal weight*, or distribution of colour and shade.

To provide effective word clouds, it is necessary to understand how the information they convey is perceived. Rivadeneira, Gruen, Muller, and Millen (2007) studied the effect of various tag cloud parameters, such as font size and word placement, to determine how they impact the types of task that users might try to accomplish. Words in largest fonts were recalled more than others, with a slight increase in recall in the top left quadrant of the word cloud compared to elsewhere (presumably with predominantly western participants). In a second experiment they found that words listed one per line in frequency order were more effective than alphabetical, or frequency-based blocks of text, but spatial clouds appeared to be the second most effective for the task of gaining a correct impression of the "tagcloud owner". For recognition tasks, where users were shown words and needed to indicate whether they were present in a word cloud shown previously, font size was the main factor influencing success.

Halvey and Keane (2007) also found that font size and word position were dominant factors in their study of word cloud use in which participants had to select a given word from a display of words. The researchers measured the time it took for participants to select the appropriate word, showing an indirect relationship between font size and time. They also observed that words found on the bottom line and the middle of clouds were selected more quickly, indicating scanning behaviour, which also occurred for alphabetical lists.

Bateman, Gutwin and Nacenta (2008) studied various tag cloud visual parameters by using a task in which participants selected the ten most important words in the cloud, based purely on visual properties. Like prior studies, font size was found to be the strongest feature, with others being font weight, intensity and colour. While colour was a factor in decisions, participants differed in which colour was chosen as the important one. However, the two colours used in the study were red and blue. If, for example, variations in colour saturation were used instead, there may have been a clearer outcome across the cohort of participants.

One issue that can impact the perception of font size is that of word length. Alexander, , Chang, Shimabukuro, Franconeri, Collins, and Gleicher (2017) found that longer words and those with letters that hang below the line, such as “g”, are perceived to be larger than words that don’t have those attributes.

In summary, we have discussed the word cloud concept, how it is used, how it has been extended, and how the various parameters of a word cloud’s construction affect their perception and usability. In the next section we discuss literature relevant to the perception of an *audio* word cloud, as well as prior work on word clouds in the audio domain.

2.2 Audio Perception and Audio Word Clouds

We introduce here some of the relevant literature regarding the perception and salience of musical notes. Francès (1958) discussed the concept of figure and ground in music perception, that is, the part of a piece of music that is likely to be considered the most salient, and therefore the melody. Typically, the melody comprises the highest notes, except in the case where high notes are long or repetitive. Deutsch (1982) described how gestalt principles applied to music, where pitch proximity is more important than good continuation and loudness, in the grouping of notes. However, the shape of a melody can also affect the perception of loudness and importance of some of its notes via melodic accents (Tekman, 1998).

Melodic accents occur when a melody note is arrived at with a “leap” of more than two semitones (Tekman, 1998), and rhythmic accents occur on notes of longer duration. Accented notes are generally perceived as more important than unaccented ones, with, for example, pitch accuracy less noticeable on unstressed notes (Sundberg, 1999). Melodically accented notes are perceived to be louder than unaccented notes (Tekman, 1997; Tekman, 1998), which would interact with their perceived relative importance. In a detailed empirically derived model of melodic accents by Thomassen (1982), and validated by Huron and Royal (1996), the relative direction of successive notes is a more important factor than the interval size for perceiving a melodic accent, with the strongest accents occurring either side of a repeated pitch, and the next strongest being a rising note which is followed by a falling note (Huron and Royal, 1996).

We located one prior publication on applying the concept of word clouds to the audio domain. Ajmera, Deshmukh, Jain, Nanavati, Rajput, and Srivastava (2012) implemented an audio word cloud in which audio documents were summarised by identifying frequently occurring words in recorded speech and rendering them as an audio word cloud. One advantage of their approach is that they didn’t use speech recognition, but instead analysed the audio directly, to enable the technique to be language independent, and to be of benefit to illiterate speakers of insufficiently resourced languages. They proposed limiting the audio word cloud to a spoken sequence of up to eight frequently occurring words. For their experiments, they varied the amplitude, “voice quality” (pitch and vocal tract size), echo and repetition of the words. They found that amplitude and voice quality were the most effective in expressing the relative importance of words in audio clouds, with voice quality being more important for low-literacy participants. For their study, the deeper the voice, the more the uttered words were perceived as important. This partially contradicts evidence from the study by Tusing and Dillard (2000), who examined the concept of dominance, as communicated via vocal quality. Amplitude (loudness) and amplitude standard deviation were positively associated with dominance, and speech rate was negatively associated with it. Pitch was only associated with dominance for male voices, in a positive relationship, with no trend for female voices. This partially agrees with Ko, Sadler & Galinsky (2015), who discuss two different types of dominance: physical and social. Their study looked specifically at social hierarchy as perceived through the voice and found that higher pitch was associated with higher status, as was higher loudness variability and reduced pitch variability. Both the above studies found that female voices were more difficult to rate for hierarchical rank, probably because participants would have more exposure to highly ranked males in society than females. As further evidence, Ko et al. (2015) noted that the factors they discovered were virtually identical to the outcome of voice training received by Thatcher to give her voice more authority.

In summary, pitch, melody shape and loudness are the main musical factors leading to a musical part being perceived as melody, whereas pitch, amplitude, loudness variability, pitch variability, and speech rate are factors that impact dominance perception of speech. In the next section we provide a brief overview of algorithmic composition and data sonification relevant to the current work.

2.3 Algorithmic Composition and Data Sonification

Algorithmic composition has a long history, with the first known instance being by Guido d'Arezzo, who systematically converted vowels of text into notes, in approximately 1026 (Edwards, 2011).

Goals and attitudes to the production of compositions algorithmically vary. Much algorithmic composition involves simulating existing styles of music. For example, the *FlowMachines* system learns musical style from a collection of musical works of a particular style, and then generates music in a similar style (Papadopoulos, Roy and Pachet, 2016). The other main branch of algorithmic composition seeks to discover new musics, discarding the conventions of existing music, including tuning systems and keys (Burt, 1996). Burt quotes Herbert Brün, who stated "We're interested in the music we don't like yet" (Burt, 2007).

A further dimension in the production of music algorithmically is the degree of intervention or post-production that occurs after the music is generated. Xenakis and Koenig were known to provide significant manual input into the final composition, whereas Hiller insisted that the algorithms should be the only source of music (Edwards, 2011).

There is considerable overlap between the concepts of algorithmic composition and data sonification, in that both attempt to automate the generation of sound based on inputs. The goals, however, are quite different. Where algorithmic composition seeks to simulate or discover new music, the goal of data sonification is to help people understand complex data through the use of sound.

Data sonification is the audio equivalent of data visualisation. Various types of data sonification have been enumerated, such as audification (converting numerical data points into values in an audio wave-form), parameter mapping (converting data into parameters of sound, such as pitch), and model-based sonification (representing a data distribution with an audio model not directly related to individual data point dimensions) (Hermann, 2002). Parameter mapping by the simple conversion of data to pitch and duration representations has been shown to communicate typical graphical information with similar effectiveness to visual graphs (Flowers, Buhman, & Turnage, 2005). As a relatively young field, the definition of sonification has been revisited periodically, with Hermann (2008) restating it as needing to be objective, systematic, reproducible and able to be used with different data. The output of sonification is expected to be sound signals, which may be triggered by a model. The definition in the Sonification

Report (Kramer et al. 2010) specifies that the sound signals should be “non-speech audio used to convey information”.

Common sonification techniques relevant to the current project include converting numerical data into notes to generate melodies. This is seen, for example, in the sonification of genetic data (Larsen, 2016), and the “songification” work of Verhoeven et. al (2014), where the two dimensions of distance from a central point and time between events was turned into pitch and duration for melody notes. Some initial work on sonification of emotion is also starting to appear (Winters & Wanderley, 2014), exploiting the arousal-valence model of emotion, in which all emotions are mapped to a numerical two-dimensional coordinate and thus can be objectively mapped further to sound in a systematic and repeatable manner.

In the current work, being a prototype algorithmic composition, there was some manual intervention to produce the final result, thus more in alignment with Xenakis than with Hiller. While songs have been produced algorithmically before, sonification using voices singing words, whether synthetic or real, seems to be a new contribution. Where the generation of sung audio is completely “objective, systematic, reproducible and able to be used with different data”, it challenges the existing definitions of sonification, which exclude speech (Kramer et al. 2010) but are silent about synthetic voices that sing. In the next section we briefly introduce relevant work related to text analysis.

2.4 Topic Modelling and Style Identification

The fields of information retrieval and natural language processing include research into how best to automatically identify the topic of text documents. This is achieved with mathematical models based on relative word frequencies. For example, the well-known TF*IDF class of formulae use a term frequency within the document in question (TF) multiplied by the inverse of the number of documents in which the term occurs (DF). (For further detail, see Zobel and Moffat (1998) for a study of TF*IDF variants for information retrieval.) This particular formula has been superseded by more complex statistical models, but at their core is the contrast of these two frequencies. Using these two components allows words such as “the” to be down-weighted, since they occur frequently in virtually all documents written in the English language, but allow topic words, such as “sonification” in this paper, to be ranked highly. The frequency of words relative to different text corpora can also reveal the *style* of writing (Brooke and Hirst, 2013). For example, writing that is personal tends to have more pronouns, such as “you”, and formal writing often uses the passive voice. The variation in vocabulary in different domains is also the basis of English language courses such as those that focus on academic English (Thurston and Candlin, 1998), and even indicate different text genres (Stamatatos, Fakotakis & Kokkinakis, 2000).

We have discussed relevant literature on the key areas relevant to the current work, being word clouds, audio perception, algorithmic composition and data sonification, and text analysis. In the next section we describe the new concept *data choralification*, and how we implemented a prototype that demonstrates the concept.

3. Data Choralification

In this work our research question was: Can word clouds be generated as choral pieces? Associated with this question was the idea that from listening to the choral word cloud, listeners would understand the topicality of the underlying set of documents. An additional element explored was the emotional content of the text. The communicative purpose of the original composition - which was part of a larger work about the aftermath of sexual assault - was to represent being overwhelmed by information on the topic of sexual assault. It was intended to be an uncomfortable corpus summary. The concept of the word cloud led to a novel method of text setting.

We used the following steps to generate the choral word cloud:

(1) Select a set of documents on a topic
(2) Calculate the word frequencies of all words occurring in the set of documents
(3) Filter out function words using a stop list
(4) Retain the most frequent k words (k set to 30)
(5) Determine the frequency range of the retained words
(6) Convert the word frequencies to audio frequencies in a defined choral vocal range (set to Bass low G (MIDI 43) to Soprano high F-sharp (MIDI 78)), using a log-based transformation that preserves the shape of the frequency range (See Figure 1)
(7) Allocate words to each of the four vocal parts, soprano, alto, tenor, and bass, such that each has a non-overlapping major sixth of note range
(8) Arrange words in frequency order within each vocal part
(9) Use quavers (eighth-notes) as the shortest duration note, and apply this duration throughout the vocal part with the most words allocated to it – usually the bass part, as there are more lower frequency words than high frequency words
(10) Arrange the shortest duration note part in a 4/4 score such that the stress of syllables coincide with strong (first beat) or medium (third beat) stressed positions in the bar
(11) For the remaining vocal parts, extend the note durations so that the parts are approximately the same duration as the short note duration part. Align word stresses to beat stresses, as with the short note duration part. Durations are either crotchets (quarter notes), minims (half notes), semibreves (whole notes) or longer in multiples of minims, with stressed syllables being extended compared to unstressed.
(12) Parts commence one beat after the other in order of first appearance in the contiguous set of documents. This results in some parts commencing with a note that is extended by a beat or two to align the stress of the next syllable with the appropriate beat of the bar.
(13) To enable this atonal work to be performed with an amateur choir, an accompanying part was created using piano, clarinet and flute to provide starting notes for each section and additional support notes where words commenced at the start of a bar.
(14) The selected tempo for the work was 60BPM, to be performed without tempo variation, aside from a slight pause between sections.

Steps 9-13 were manual for the generation of the prototype work. The automated parts of the process consisted of a set of scripts in python and bash, making use of various unix utilities. Calculation of the MIDI note values for the prototype was completed in a spreadsheet.

The calculation of the MIDI note number was as follows:

$$\text{midinumber} = \text{round}(69 + 12 * \mathit{log}_2(\text{audiofreq}/440)\ )$$

where:

$$\text{audiofreq} = \text{maxaudio} * (\text{wordfreq}/\text{maxword})^{\text{log\_const}}$$

and:

$$\text{log\_const} = 1/\mathit{log}_2(\text{maxword}/\text{minword})^{(1/(\ \mathit{log}_2(\text{maxaudio}/\text{minaudio})))}))$$

where minaudio is the audio frequency of Bass low G (~98Hz, or MIDI note number G2), maxaudio is the audio frequency of Soprano high F# (~734Hz, or MIDI note number F#5); wordfreq is the frequency of the word; maxword and minword are the maximum and minimum word frequencies respectively of the selected range of words. The effect of the above formula is that the notes assigned are spread precisely over the selected MIDI range, and preserve the general shape of the distribution. This is demonstrated in Figure 1, which graphs the rank and frequency of both words and audio, for the small "letters" data-set and the full data-set. Regardless of the number of words retained for the word cloud, the lowest frequency word, and therefore note, is rendered as Bass G2, and the highest frequency word and note, as Soprano F#5.

4. Application

The prototype choralification was applied to sets of documents on the topic of sexual assault. The first set of documents consisted of on-line resources for victims/survivors of sexual assault. The second was a set of letters found and downloaded using the search query "letter to my rapist" – as part of the therapy for recovery from sexual assault, victims/survivors are encouraged to write a letter to their rapist/assailant. The third was a set of legal documents regarding sexual assault. In addition, the full set of documents was used together. Each of the document sets were processed to produce high frequency word sets with the number of words set to 30. As a coda, the top 6 words of the combined set were processed.

The work was submitted as part of a larger work to be performed at a concert of new choral works. Performance occurred after a semester of weekly rehearsals of a large repertoire.

4.1 Performance Preparation

The atonal nature of the work was manageable for the choir with the instrumental accompaniment as support. However, the conductor elected to provide full rehearsal piano support during the performance.

Due to the triggering topic, rehearsal was difficult for sopranos and altos, as a large percentage of women have experienced sexual assault or are adversely affected by the theme. Trigger warnings were provided on the score, with explicit permission for choristers to withdraw from rehearsal and performance if they chose. Some of the worst affected singers chose to persevere and perform the work. An additional women-only rehearsal was scheduled, as well as rehearsal without words, to assist in preparation.

4.2 Evaluation

Evaluation of the work consists of feedback from an expert panel, informal feedback, and reflection. The choralification was submitted to a choral composition competition as one movement of a larger work. Due to length and resource constraints on eligible works, the entire work was submitted as two works of three movements each. The expert panel consisted of three composers, two of whom are accepted members of the national body of composers, for which membership requires having won awards, commissions or recording contracts for composition. The panel members were not experts in algorithmic composition or sonification and were judging works based on their quality as choral compositions suitable for the choir hosting the competition.

4.3 Results

Table 1 shows the source text size, the filtered number of words, and the number of words allocated to each part. (Note that one word was omitted from the bass part for the “everything” text, so it would have had 22 words if retained.) Applying the data choralification as described led to many words (up to 21) being in the bass vocal part, and on one occasion only one word in the soprano part. There was, on average across the data sets, a monotonic increase in the number of words per part, from soprano to bass. It is to be expected with word frequency distributions (Baayen, 2001), that as the word frequency decreases, more words will be found at that frequency, and that the gaps between frequencies will become smaller. As a consequence, the size of the intervals between consecutive notes also becomes smaller with decreasing word frequency. This can be seen in the excerpt shown in Figure 2. The soprano part has a leap of a perfect 4th, the alto has a major 3rd, the tenor tends to move in tones, while the bass is either static or moves in semitones.

The distribution of words is somewhat dependent on the size of the text collection being choralified, as well as the number of words to be included in the word cloud. For example, where the combined collection is used, the threshold of 30 words leads to many words being in the bass part, whereas with a threshold of six words, the bass part has a single word, and the tenor and alto parts two words each. While the general trend for the collections is an increase in the number of words as the frequency drops, it is not consistently so for any particular collection and threshold. For example, the Laws collection is the only one that has a strictly increasing number of words per part. It is more likely when the *type-token ratio*, that is, the number of distinct words divided by the total number of words, is low.

An artefact of the simple process of filtering words caused some words with apostrophes to be retained that otherwise might have been excluded, such as “don’t”.

Expert panel review of the work focused on the instrumentation: “could be employed more interestingly”. Additional feedback from a panel member indicated that “the important words in the bass line felt a bit buried in the texture”.

Informal feedback from choristers was that it was “powerful”, and one singer was profoundly and negatively affected by this piece due to the male voices singing “those words”.

5. Discussion

5.1 Words and Word Frequency

The main research question being addressed by this prototype was whether word clouds could be generated as choral pieces. This choral work demonstrates that it is possible to create a choral work inspired by word clouds, that shares some of their features. The frequency of occurrence of words is clearly demonstrated by pitch and vocal part due to it being a direct mapping. However, due to the polyphonic nature of the work, and the soprano range overlapping the frequency range of vowels (see for example, Deme, 2014), it is highly likely that not all words were clearly understood by the audience, but even if they only understood, say, twenty percent of the words sung there is a good chance that the topic would be perceived, based on a random sampling of words. While intelligibility of sung text has been discussed since at least the sixteenth century (Monson, 2002), there appears to be little research to precisely determine the circumstances under which it is understood, so our estimates here are guesswork.

Due to the nature of attention in music perception, the lower frequency words were perceived by at least one expert listener as more important than the high frequency ones, despite the criticism that they were not very clear. Whether they are indeed the more important words is uncertain. In general, highly frequent words tend to be function words such as “the” and are clearly not considered important in terms of the topic of the text. These were excluded from the composition for this reason. Highly frequent words within the text but not in general (in this example “sexual”) tend to indicate the broad topic of the document. The slightly lower frequency but still quite prevalent words indicate the way the topic is being discussed. All the example texts are about sexual assault and have either that term, “rape”, or both, as frequent terms, but differ in the other frequently occurring terms. For example, in the first set of documents there are more words such as “support”, “resistance”, “survivor”; the second has more personal words, such as “remember”, “pain”, “feel”; while the third has legal words “evidence”, “law”, “legislation”. For the example work, this difference is a feature that provides textual contrast between sections. For communicating the nature of the document sets the composition is successful – again on the assumption that a reasonable percentage of the words were understood.

As was shown in Table 1, there are usually few words at the highest frequency range. While pitch was the main variable used to represent importance, a side effect of the power law relationship of word frequency distributions led to typically long durations for these notes in this sonification. Consequently, while font size makes the highest frequency word the most prominent in the image in a visual word cloud (Rivadeneira et al., 2007; Halvey & Keane, 2007; Bateman et al, 2008), using pitch and extended duration doesn’t necessarily do this, since movement in another part tends to override a static highest pitch part. However, whenever the pitch changes in the highest vocal parts, it is with a leap, thus creating a melodic accent, which may be more salient than the stepwise movement of the lower parts (Tekman, 1997; Tekman, 1998). On the whole, in terms of musical salience, there would have been a shift in perceived importance whenever a new high note word was sung, with attention then shifting to the lower moving parts, once the part becomes static. This may prevent the whole word being preceived, since each syllable was rendered at the same pitch, and attention would shift before the first syllable is completely sung.

In terms of perceived dominance, the bass part could be perceived as being of high physical dominance (Ko et al., 2015). The features associated with *social* dominance are high (male) pitch, low pitch variability, slow speech and high loudness variation. The pitch cue only applies to male voices. If the tenor part had lower pitch variability, then it would generally be perceived as higher in social dominance than the bass part, but the high pitch and slower speed are accompanied by more pitch variability in this composition, sending a mixed signal. Slow speed would tend to predict the soprano part as dominant. Thus, all parts except perhaps the alto line have at least one feature of dominance. None of the parts employed loudness variation – a factor of social dominance.

A potential alternative method of ensuring that high frequency words are perceived as the important ones – if that is intended – is to have them repeated by the vocal part rather than extended. This may, however, make all the words sung less clear, and repetition of notes also reduces perceived importance (Francès, 1958), so it is unlikely to be helpful. It was also found to be less effective for the spoken audio word clouds studied by Ajmera et al. (2012). Another alternative is to have an inverse relationship between word frequency and note frequency, so that the typically sustained notes from high frequency words occur in the bass part while shorter notes associated with less frequent words are sung by sopranos. A third possibility is to not have a direct relationship between the word frequency and audio frequency, thus allocating words more evenly between parts, but retaining the relative ordering of frequency. This would result in a loss of information about the relative magnitude of the word frequencies but is probably a closer match to what happens with visual word clouds and font size. However, perhaps the most promising feature to add to the part *intended* to be perceived as the most important is accents, thus providing loudness variation, the currently unused factor identified with social dominance, potentially improving the clarity of those highlighted words.

5.2 Tonality

In word clouds created via the Wordle program, users are able to set various parameters to change the appearance of the word cloud, including font, colour, and layout. These are aesthetic choices of the user that may allow them to convey something about the text that they have selected. In a similar manner, the present composition used musical parameters to convey the nature of the content.

The mapping chosen was atonal, which is often perceived as uncomfortable compared to consonant music that conforms to cultural traditions of the listener. This was a conscious choice due to the nature of the text being set. Based on informal feedback received, this aspect of the sonification was successful. It remains to be seen how a mapping that can communicate other sentiments can be created. The options include constraining the notes to a diatonic scale, presenting the words in a different order to create melodies rather than descending passages, and varying the articulation of the sung words.

5.3 Automation and Future Application

The atonal music was challenging but not impossible for singers to sing. However, the concepts demonstrated by this algorithmic choral composition can potentially be extended to create a useful sonification of text documents, in which the audio is generated automatically. To that aim we have rendered the audio using synthetic voices, via the Sinsy[3] singing synthesizer. The clarity is variable, but it may be sufficient to communicate the content, and singing synthesis models should improve over time.

[3] http://sinsy.jp

One of the issues with polyphonic rendering is that it can make words harder to comprehend, as was raised at the Council of Trent when discussing liturgical music in 1562 (Monson,2002). For the prototype sonification application, we have automated the generation of a monophonic descending sequence of sung words to represent the frequent words of a set of documents. This retains the word frequency information and maximum clarity, and coupled with its automation, makes it a potentially useful summary of documents, possibly also for the vision-impaired or the illiterate (Ajmera et al., 2012). It also matches the most usable visible text display, as discovered by Rivadaneira et al. (2007). For the example shown in Figure 3, taken from Jane Austen's *Sense and Sensibility*, the same vocal range was used as for the choralification.

Most screen readers for the vision-impaired allow the speed of the spoken text to be selected by the user for maximum efficiency of use. Therefore we suggest that a rendered word-frequency-based document summary with tempo control can provide an additional dimension via pitch, while still being similar to a currently familiar mode of receiving information for this group of users. The utility and usability of this approach is reserved for future research.

6. Conclusions and Future Work

Much as the creator of Wordle was inspired by Matt Jones and others to implement an aesthetic rendering of tag clouds, we were inspired by Wordle to create a choral equivalent of word clouds (and later an automated sonification-based text summary). We created a data choralification in which the most frequent words (excluding stop words) were given an audio frequency that was directly related to their frequency within the text, and the words distributed amongst four vocal parts that sang their words in descending frequency order. The choral composition was inspired by the concept of word clouds by rendering frequent words with some perceivable method of representing their relative frequency. However, font size tends to be perceived as representing the relative importance of words within a word cloud, whereas the present choralification doesn't have this effect. Given that different frequency ranges in a document can provide different glimpses of it, and may be equally important for some applications, the choralification may represent this ambiguity better than visual word clouds do. A possible extension could intentionally separate specific aspects of documents such as topic, affect and style through more nuanced grouping of words prior to rendering in a score.

This work provided a novel approach to text setting, in that it produced a word frequency-based corpus summary of the source text, rather than setting the text as a whole, or recognisable excerpts from it. The area of algorithmic text setting or pre-processing prior to setting is underexplored and warrants further research.

In the present work we used a simple filter of stop words applied to the text. There are other alternative ways of ranking the importance of words in a document or set of documents that may be more effective at choosing topic words, including alternative stop lists and formulae based on term and document frequency. Also, in the example set, it may have been more relevant to include some of the filtered words such as "you", which contrasted with the more formal texts on the topic and made them more immediate when sung. Such frequent words can be selected based on their perceived affect, which can be crowdsourced by capturing relative affect on the three dimensions of emotion: arousal, valence and dominance (Mohammad, 2018).

Another possible extension is to use topic modelling to separate the words into different topic classes before rendering as sung notes, leading to a choral equivalent of the work by Paulovich et al (2012).

In this example work, the music was rendered as an atonal set of descending vocal lines, which matched the uncomfortable nature of the topic. However, it would be interesting to combine sentiment analysis of texts with different rendering techniques, so that the emotion of the text can be represented, in addition to the topic words. This is an area we are exploring, and that is relatively underexplored in data sonification (Winters & Wanderley, 2014).

The work also led to questions regarding the intelligibility of sung words at different pitches and in a polyphonic texture, as well as the perception of the relative importance of words sung at different speeds and pitches. Experiments with listeners would be needed to determine the answers.

The composition was successfully performed as one movement of a larger work in a choral concert. However, future work will involve extending the exploration of automatically generated audio via synthetic voices to provide usable audio text summaries.

References

Ajmera, J., Deshmukh, O. D., Jain, A., Nanavati, A. A., Rajput, N., & Srivastava, S. (2012, February). Audio cloud: creation and rendering. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (pp. 277-280)*. ACM.

Alexander, E. C., Chang, C. C., Shimabukuro, M., Franconeri, S., Collins, C., & Gleicher, M. (2017). Perceptual Biases in Font Size as a Data Encoding. *IEEE transactions on visualization and computer graphics*.

Baayen, R. H. (2001). *Word frequency distributions* Text, Speech and Language Technology (Vol. 18). Springer Science & Business Media.

Bateman, S., Gutwin, C., & Nacenta, M. (2008, June). Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the nineteenth ACM Conference on Hypertext and Hypermedia* (pp. 193-202). ACM.

Brooke, J., & Hirst, G. (2013, October). Hybrid Models for Lexical Acquisition of Correlated Styles. In *IJCNLP* (pp. 82-90).

Burt, W. (1996). Some parentheses around algorithmic composition. *Organised Sound*, 1(3), 167-172.

Burt, W. (2007). Some Musical and Sociological Aspects of Australian Experimental Music. *Resonate Magazine*, 31, originally published in Sounds Australian, 1993.

Chowdhury, G., Koya, K., & Philipson, P. (2016). Measuring the impact of research: lessons from the UK's Research Excellence Framework 2014. PloS one, 11(6), e0156978.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X., & Qu, H. (2010, March). Context preserving dynamic word cloud visualization. In Visualization Symposium (PacificVis), 2010 IEEE Pacific (pp. 121-128). IEEE.

Deme, A. (2014). Formant strategies of professional female singers at high fundamental frequencies. In Proceedings of the 10th International Seminar on Speech Production (ISSP) (pp. 90-93).

Deutsch, D. (1982). Grouping mechanisms in music. In The Psychology of Music (pp. 99-134).

Edwards, M. (2011). Algorithmic composition: computational thinking in music. *Communications of the ACM*, 54(7), 58-67.

Feinberg, J. (2010). Wordle. In Steel and Iliinsky (2010), 37--58.

Flowers, J. H., Buhman, D. C., & Turnage, K. D. (2005). Data sonification from the desktop: Should sound be part of standard data analysis software? *ACM Transactions on Applied Perception* (TAP), 2(4), 467-472.

Halvey, M. J., & Keane, M. T. (2007, May). An assessment of tag presentation techniques. In Proceedings of the 16th international conference on World Wide Web (pp. 1313-1314). ACM.

Hearst, M. (1999). User interfaces and visualization. *Modern information retrieval*, 257-323.

Hearst, M. A., & Rosner, D. (2008, January). Tag clouds: Data analysis tool or social signaller?. In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual (pp. 160-160). IEEE.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word cloud explorer: Text analytics based on word clouds. In System Sciences (HICSS), 2014 47th Hawaii International Conference on (pp. 1833-1842). IEEE.

Hermann, T. (2002). Sonification for exploratory data analysis (Doctoral dissertation).

Hermann, T. (2008). Taxonomy and definitions for sonification and auditory display, Proc. ICAD 2008, IRCAM, France.

Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. Music Perception: An Interdisciplinary Journal, 13(4), 489-516.

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2015). The sound of power: Conveying and detecting hierarchical rank through voice. *Psychological Science*, 26(1), 3-14.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J. H., Miner, N., & Neuhoff, J. (2010). Sonification report: Status of the field and research agenda.

Larsen, P. E. (2016). More of an art than a science: Using microbial DNA sequences to compose music. *Journal of microbiology & biology education*, *17*(1), 129.

Lohmann, S., Heimerl, F., Bopp, F., Burch, M., & Ertl, T. (2015, July). Concentri cloud: Word cloud visualization for multiple text documents. In *Information Visualisation* (iV), 2015 19th International Conference on (pp. 114-120). IEEE.

Marszałkowski, J., Rusiecki, Ł., Drozdowski, M., & Narożny, H. (2014, August). Toward building aesthetic, useful and readable tag clouds for websites. In *e-Business* (ICE-B), 2014 11th International Conference on (pp. 230-235). IEEE.

McMillan, K. K., Pugh, M. J., Hamid, H., Salinsky, M., Pugh, J., Noël, P. H., ... & LaFrance, W. C. (2014). Providers' perspectives on treating psychogenic nonepileptic seizures: frustration and hope. *Epilepsy & Behavior*, 37, 276-281.

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The qualitative report*, 15(3), 630.

Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 174-184).

Monson, C. A. (2002). The Council of Trent Revisited. Journal of the American Musicological Society, 55(1), 1-37.

Papadopoulos, G., & Wiggins, G. (1999, April). AI methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity* (pp. 110-117). Edinburgh, UK.

Papadopoulos, A., Roy, P., & Pachet, F. (2016, September). Assisted Lead Sheet Composition using FlowComposer. In *International Conference on Principles and Practice of Constraint Programming* (pp. 769-785). Springer International Publishing.

Paulovich, F. V., Toledo, F. M., Telles, G. P., Minghim, R., & Nonato, L. G. (2012, June). Semantic wordification of document collections. In Computer Graphics Forum (Vol. 31, No. 3pt3, pp. 1145-1153). Oxford, UK: Blackwell Publishing Ltd.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007, April). Getting our head in the clouds: toward evaluation studies of tagclouds. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 995-998). ACM.

Rowell, A., Evans-Reeves, K., & Gilmore, A. B. (2014). Tobacco industry manipulation of data on and press coverage of the illicit tobacco trade in the UK. *Tobacco control*, tobaccocontrol-2013.

Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2000, July). Text genre detection using common word frequencies. In Proceedings of the 18th conference on Computational linguistics-Volume 2 (pp. 808-814). Association for Computational Linguistics.

Steele, J., & Iliinsky, N. (2010). Beautiful visualization: looking at data through the eyes of experts. O'Reilly Media, Inc.

Sundberg, J. (1999). Perception of singing. *The psychology of music*, 3, 69-105.

Tekman, H. G. (1997). Interactions of perceived intensity, duration, and pitch in pure tone sequences. *Music Perception: An Interdisciplinary Journal*, 14(3), 281-294.

Tekman, H. G. (1998). Effects of melodic accents on perception of intensity. *Music Perception: An Interdisciplinary Journal*, 15(4), 391-401.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for specific purposes*, 17(3), 267-280.

Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. The Journal of the Acoustical Society of America, 71(6), 1596-1605.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. *Human Communication Research*, 26(1), 148-171.

Verhoeven, D. K., Davidson, A., Gionfriddo, A., Verhoeven, J., & Gravestock, P. (2014). Turning Gigabytes into Gigs:“Songification” and Live Music Data. *Academic quarter*, 9, 151-163.

Viegas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualization with wordle. *IEEE transactions on visualization and computer graphics*, 15(6).

Winters, R. M., & Wanderley, M. M. (2014). Sonification of Emotion: Strategies and results from the intersection with music. *Organised Sound*, 19(1), 60-69.

Zobel, J., & Moffat, A. (1998, April). Exploring the similarity space. In *ACM SIGIR Forum* (Vol. 32, No. 1, pp. 18-34). ACM.

Tables

Table 1. Source Text Distribution

Source Text (Size in (Words)	Distinct Words (Filtered Words)	Type/Token Ratio	Sop Words	Alto Words	Tenor Words	Bass Words
Resources (16,047)	3,033 (2,741)	18.9%	3	2	7	18
Letters (16,226)	2,568 (2,295)	15.8%	4	3	12	11
Laws (13,292)	1,744 (1,501)	13.1%	1	4	7	18
Everything (44,559)	5,299 (4,955)	11.9%	2	3	3	21
Coda			1	2	2	1

Note*: The table shows descriptive statistics of the document collections used to produce the prototype, and the number of words allocated to each choral voice part, given the chosen method of converting word frequencies to audio frequencies. As can be expected, fewer words are at the very high frequencies, and more in the bass range.*

Figures

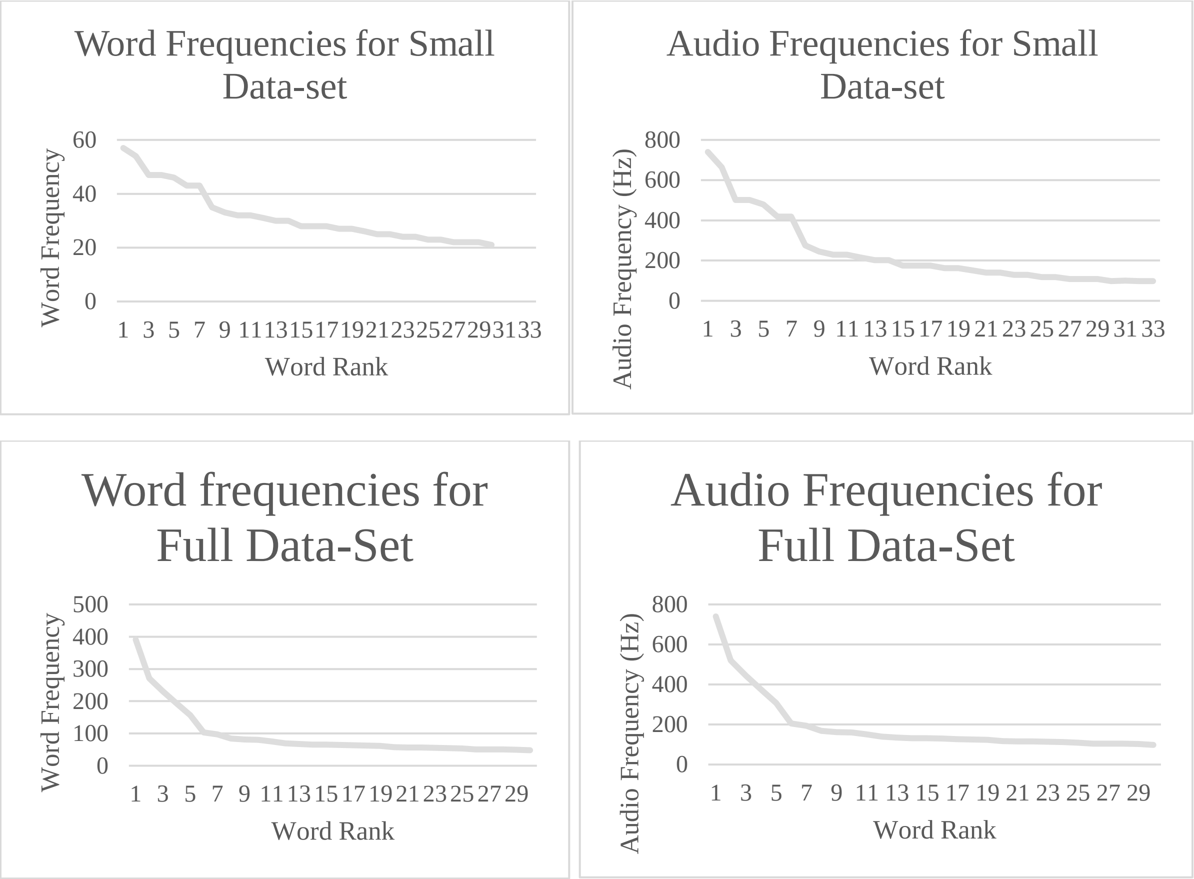

Figure 1. Graphs showing the rank-frequency distribution of text and musical notes.

Figure 2. Excerpt of World Cloud, a prototype data choralification.

Figure 3. Prototype word frequency sonification of Jane Austen’s Sense and Sensibility.