[[1]](#footnote-2)

Grammar Informed Sound Effect Retrieval for Soundscape Generation

Emmanouil Theofanis Chourdakis and Joshua D. Reiss

Queen Mary University of London, [e.t.chourdakis@qmul.ac.uk](mailto:e.t.chourdakis@qmul.ac.uk)

*Abstract*— This paper introduces a simple method for retrieval of sound effects from a sound library, with the goal of creating soundscapes from a sentence in natural language. Grammatical constituents are extracted from the sentence in order to construct search queries relevant to the objects described in the sentence. A comparison against a previous method of retrieving sound effects for soundscape generation is shown.

# Introduction

An essential part of soundscape generation from text is retrieving relevant sound effects. Consider sentence (1): *Crows are feeding on rubbish at a garbage dump.* We expect to retrieve sound effects relevant to crows and the noise they make while feeding, and the sounds of a garbage dump. Current methods that generate soundscapes from text do retrieval by constructing text queries from generic sentence features. Such approaches do not take in consideration the role of the words in a sentence and can lead to undesirable results. For example, when using n-grams (e.g. in [1]) word combinations such as f*eeding rubbish* will be usedto construct queries. This might lead to retrieve *feeding rubbish to a cat* which we expect to be different from *crows feeding on rubbish*. In this paper we aim to constrain searches to sound effects relevant to the *subject* and *object* of the sentence in order to avoid such problems. Main motivation for this paper is previous work on producing sound from story narrative where we needed methods to convey story text as sound [2].

# Approach

We suggest before constructing queries to use a simple rule-based information extraction approach, as in [3], to do sentence simplification and extract the *Subject (S)*, *Verb (V), Object (O),* and *Adverbials (A)* of the simplified sentences. We then combine those parts to construct queries. The main assumption is that a sentence describes one or more events and each event is an action of something (*subject*) possibly interacting with something else (*object)* which happens somewhere (*prepositional adverbials)*. Simplification transforms the original sentence to simpler sentences where each describes a single event. An initial query is constructed by concatenating all the parts while omitting auxiliary verbs and articles for each sentence. Since the relevant sound effects might not be available in the library for the complete query, we gradually omit those parts that are less important. For example, in (1)the sound of crows feeding on rubbish may not be available but the sound of crows feeding on something else might exist and will be more relevant than e.g. *cats* feeding on rubbish. Finally, if the *adverbial* contains a preposition such as *at*, we assume that it tells us where the action happens (*at a garbage dump, on the road,* etc.) Taking sentence (1) as an example, our method first simplifies it to sentences (2a): *Crows are feeding on rubbish* and (2b): *Crows are feeding at a garbage dump where* *S*2a,2b *= Crows, V*2a,2b *= Feeding, A*2a *= On rubbish,* and *A*2b *= At a garbage dump.* It then constructs a query by concatenating the *S, V,* and *A* parts together. Additional queries are constructed by removing *A,* then *V*. Since *A* contains a preposition (*at)* the part that follows the preposition (*garbage dump*) is also added. The queries we construct as well as queries constructed using [1] can be seen in Table 1.

Table 1. queries by proposed approach and [1] for sentence (1)

| Proposed Approach | Thorogood et al [1]. |
| --- | --- |
| Crows feeding on rubbish at garbage dump, Crows feeding on rubbish, Crows feeding at garbage dump, Crows feeding, Crows, Rubbish, Garbage dump | Crows feeding rubbish garbage dump, crows feeding rubbish garbage, feeding rubbish garbage dump, crows feeding rubbish, …, crows, feeding, rubbish, garbage, dump |

# Conclusion

We presented a method that given a source sentence can construct queries to retrieve sound effects relevant to the objects described in that sentence. While this method only works on sentences that have grammatical structure, it can retrieve relevant sound effects for sentences describing events, such as ones found in stories, supplementing our work done in [2] as a radio-play generation system from story text.

# References

1. Thorogood, M., Pasquier, P., Eigenfeldt, A., “Audio Metaphor: Audio information retrieval for SoundScape Composition”, Sound *and Music Computing, 2012.*
2. Chourdakis, E.T., Reiss J.D., “From my pen to your ears: Automatic production of radio plays from unstructured story text”, *Sound and Music Computing, 2018.*
3. Corro, L, Gemulla, R., “ClausIE: Clause-Based Open Information Extraction”, *Intl. Conf. World Wide Web, 2013.*

1. [↑](#footnote-ref-2)