Introduction:

The FOAMS sound bank was developed to specifically curate a variety of sound stimuli for misophonia research. Individual sounds from freesound.org were initially compiled based on previous literature and systematically selected for use in the present study. The individually selected audio files were further categorized into classes for which each sound occurrence was labeled according to a taxonomy. A final subset of labeled audio files were segmented and processed to create stimuli for research.

Audio Labeling:

Sound Search:

All initial potential sounds were compiled using the freesound.org API and organized by term (Figure 1). Terms were taken from previous research and prioritized based on how well they correlated with misophonia severity (Hansen et al., 2021). Each term was submitted as a search query to freesound.org until there were no more results or 1000 results were reached for the term. Results were further prioritized based on whether they met the following criteria: a creative commons zero license, a sampling rate of at least 44100 Hz, a total duration of between four and 150 seconds, and being the first audio file uploaded by a unique Freesound user (after sorting by priority).

Next, sounds from each term were played in order of priority. The "current_term_present" and "other_term_present" columns were populated using "yes", "no", or "unsure", depending on whether there was a clear instance of the selected term or another term on the list, respectively. The "notes" column was populated with any other present sounds and miscellaneous considerations about the file's suitability for the database. The "background_noise_present" column was populated using "yes", "no", or "unsure", depending on whether or not the annotator subjectively perceived the overall audio file as noisy.

This process was repeated for each sound in each of the first 10 terms, until 5 instances were reached with "yes" in the "current_term_present" column and "no" in the "background_noise_present" column. Some extra categories were populated in anticipation of expanding this sound bank past 10 terms. These extra categories were considered in the initial sound search but ultimately not annotated in the current version.

term	url	priority	current_term_present	other_term_present	notes	background_noise_present
sipping hot liquid	https://freesound.org/people/Kinoton/sounds/462270/	1	Yes	Yes	Others: swallowing	No
slurping	https://freesound.org/people/launchsite/sounds/556780/	2	Yes	No	Straw sound	No
slurping	https://freesound.org/people/ASMR_Tingles/sounds/530350/	3	Yes	Yes	Others: swallowing	No
slurping	https://freesound.org/people/charliemidi/sounds/68145/	4	Yes	Yes	Others: swallowing, footsteps, murmur	No
slurping	https://freesound.org/people/redcheek/sounds/251984/	5	Yes	Yes	Others: exhaling, taping mug on table	No
slurping	https://freesound.org/people/bfederi1/sounds/382029/	6	Yes	Yes	Others: plastic tapping	No

Figure 1. Example of sound search, depicting columns used for initial organization.

Annotations:

Each sound file (e.g., Figure 2A) was labeled using the Audacity software (v3.2.0). First, following the structure of the UrbanSound8k dataset (Salamon et al., 2014), each auditory object comprising the sound was subjectively labeled for salience by indicating whether the instance

was perceived to be in the foreground ("C1") or background ("C2") of the recording. Following the saliency label, the instance was identified using a sound class label. For example, "C1-chewing_gum" would describe a portion of sound identified as chewing gum in the foreground of the recording. Auditory objects were manually labeled within 50ms of the onset and offset of the instance (see Figure 2B). Label files were then saved and exported, specifying the sound ID number in the file name (e.g., "272420_labels.txt"). This process was repeated for each audio file selected from the previous "sound search" step.

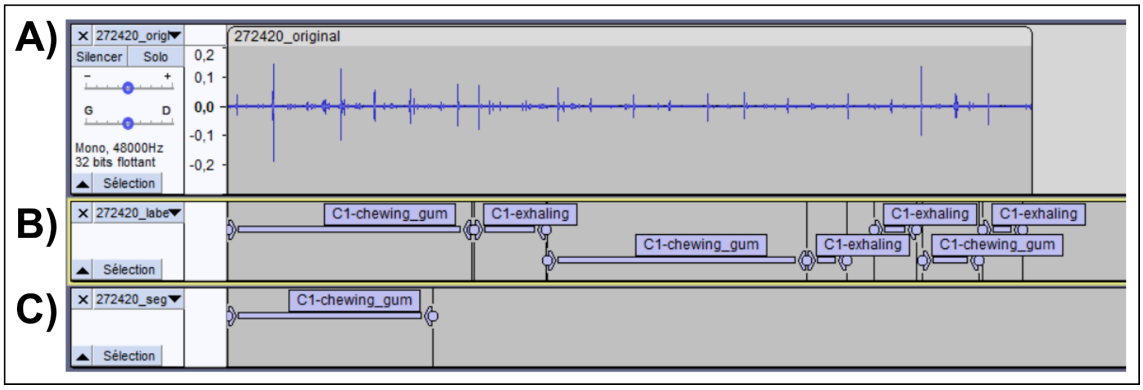

Figure 2. Annotations performed in the Audacity software. A) Original audio file from freesound.org. B) Labels describing salience and class of each auditory object present in the audio file. C) Single segmentation within the audio file most representative of the sound class.

Five instances of each of the final ten categories were further segmented into stimuli ranging from three to seven seconds. The timestamps specifying the onset and offset of the segments were annotated in Audacity (Figure 2C) and exported as text files, named according to the sound ID (e.g., "272420_segment.txt"). The original audio files were cropped according to these segmentation labels, then preprocessed (e.g., resampled to 44100 Hz). All the exported annotation text files from the initial labeling and segmentation steps are available in an associated GitHub repository, in the labels and segmentation directories, respectively.

Taxonomy:

Taxonomic categorization of sounds has been studied in the context of general acoustic soundscapes and urban environmental sounds (e.g., UrbanSound8k; Salamon et al., 2014). However, the ways in which misophonia stimuli should be categorized is often implicitly assumed and remains an open research question. Therefore, the taxonomy presented here (Figure 3A) and available on GitHub was created by subjectively choosing higher-level semantic classes based on a combination of those used in UrbanSound8k and trigger subdivisions used in previous misophonia research. Then, lower-level sound classes were dynamically added corresponding to auditory objects as they were identified in the annotation step.

Each sound class used to annotate the audio files in Audacity was added to a JSON file describing the taxonomy. Each entry included a machine-readable version of the sound class

(e.g., “chewing_gum”), a label for the visual representation of the taxonomy (e.g., “chewing gum”), and a parent category describing the source of the sound (e.g., “oral_nasal”). See Figure 3B for examples of taxonomic entries.

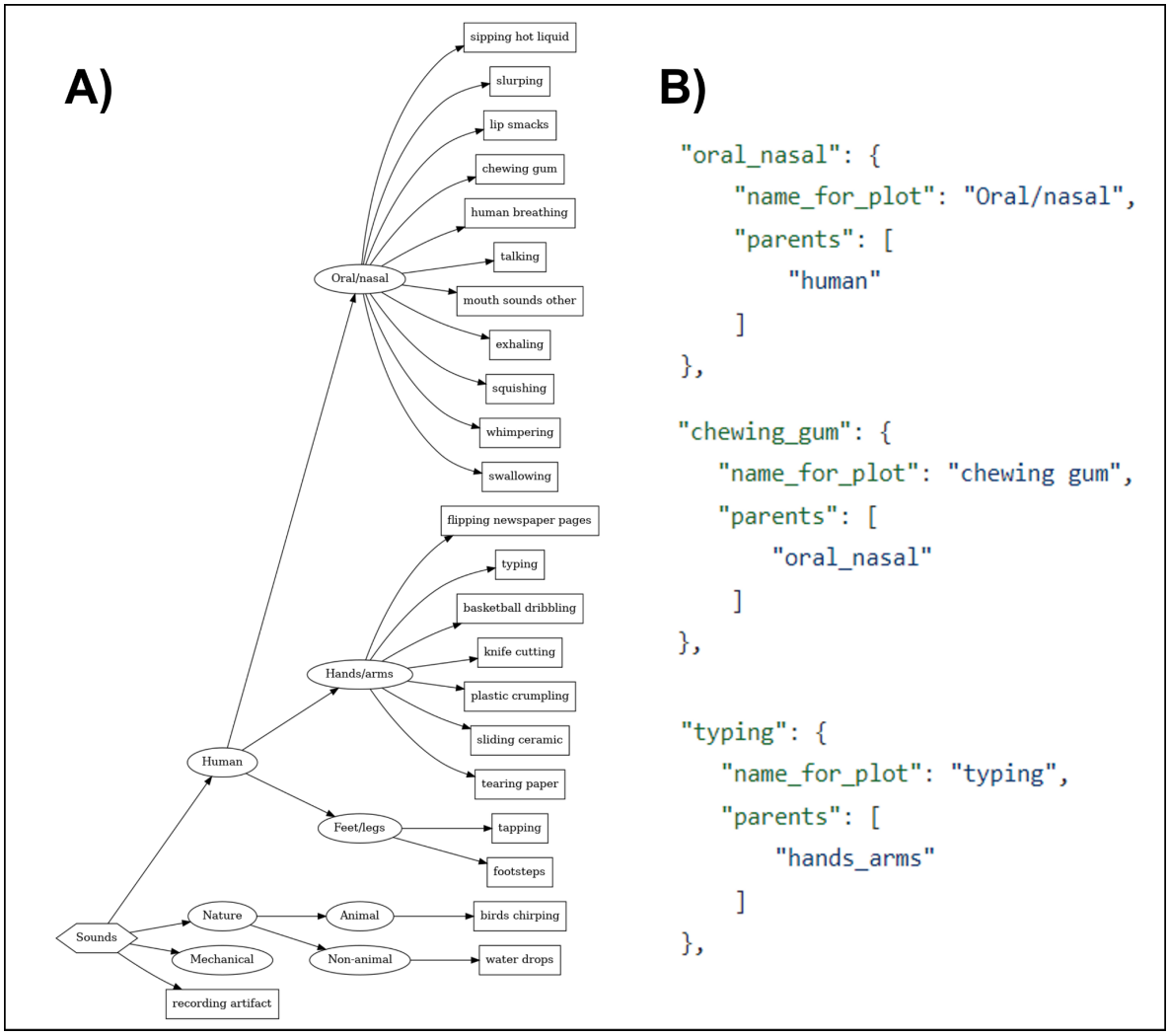

Figure 3. Taxonomic categorization. A) Plot of entire taxonomy derived from annotating 10 classes of sounds. Arrows point from parent category to child class name. B) Example entries in the JSON file to construct the taxonomy.

References:

Hansen, H. A., Leber, A. B., & Saygin, Z. M. (2021). What sound sources trigger misophonia? Not just chewing and breathing. *Journal of clinical psychology*, *77*(11), 2609-2625.

Salamon, J., Jacoby, C., & Bello, J. P. (2014, November). A dataset and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1041-1044).