# Show, Not Tell: A Pattern-Based, Deaf-Centric Classification Approach for Everyday Sounds

Anonymized

Current sound recognition systems for deaf or hard of hearing (DHH) people identify sound sources (*e.g.,* dishwasher) or discrete events (*e.g.,* door knocks). However, since different sources can produce similar sounds (*e.g.,* both washing machine and dishwasher produce a "hum"), the categories often overlap. We introduce a novel approach to categorizing sounds based on their underlying sound patterns. To ensure our classification approach aligns with the Deaf-centered practices of describing sounds, we interviewed eight ASL interpreters on how they sign different sounds. Through cluster analysis of the interpreter responses, we arrived at an 18-category taxonomy that distinguishes sound patterns based on their ASL descriptions. We evaluated our taxonomy with nine DHH people, finding the initial promise of inferring sound events and other sound cues from the signing patterns. We also used our taxonomy to train a sound recognition model, revealing a near-perfect classification accuracy on a small dataset.

CCS CONCEPTS • Human-centered computing • Accessibility • Empirical studies in accessibility

**Additional Keywords and Phrases:** Accessibility, deaf and hard of hearing, sound awareness, taxonomy

**ACM Reference Format:**

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

## 1 INTRODUCTION

Several past studies have shown that people who are deaf or hard of hearing (DHH) want increased sound awareness to access safety-related cues, perform everyday tasks, and generally be aware of their environment [2,6]. Motivated by this challenge, researchers have built sound recognition solutions that identify sound sources (*e.g.,* dishwasher, microwave) or discrete sound events (*e.g.,* door knocks) from audio signals [17,18]. However, since different sources or events can produce similar sounds (*e.g.,* both washing machine and dishwasher can produce a "mechanical humming" sound), the source or event-based categories often overlap with each other. Indeed, in a recent field study of a mobile sound recognition system, multiple similar sounding sounds were confused (*e.g.,* microwave beep and EKGs, alarm clock and phone ring), leading to poor experience for the DHH end-users [14]. This limitation inspired us to investigate an alternative sound classification scheme that classifies sounds based on the underlying patterns of sounds. We hypothesized that using the information enclosed in the sound patterns in combination with the knowledge of their situated contexts, DHH people will be able to make sense of the sound events and increase their sound awareness.

We began by compiling a list of sound patterns that covered DHH people's desired sounds. To ensure that these sound patterns can be understood by DHH people, we interviewed eight sign language interpreters to describe how they may sign these patterns to DHH people in ASL. Based on the uncovered insights, and through discussions with our multidisciplinary research team of one DHH researcher, three hearing researchers, two CART writers, and two ASL interpreters, we articulated an 18-category taxonomy that classifies sound patterns based on their ASL descriptions.

We then conducted two preliminary evaluations of our taxonomy to investigate whether (1) DHH participants could recognize the desired sound events across different contexts based on the sound patterns from our taxonomy and (2) the pattern-based classes delineated by the taxonomy could be recognized and classified algorithmically. Results from the first study with nine DHH participants provided initial evidence that DHH people could identify a wide range of sound events (*e.g.,* nail on the chalkboard and book drops onto the ground) that could occur in different environments when being presented with the sound pattern descriptions in our taxonomy (*e.g., screech/squeak* and *thump*). Moreover, participants' responses indicated that our pattern-based classification approach allows them to distinguish different sounds produced by the same object (*e.g.,* door slam vs. door knock) and to recognize similar sounds produced by different sound sources (*e.g.,* blender in the kitchen vs. coffee grinder in the coffee shop). For the second evaluation study, we used our taxonomy to train a sound recognition model to ensure that the sound patterns demonstrated by the taxonomy could be recognized and distinguished through algorithmic approaches. We found a near-perfect classification accuracy (98.6%) when evaluated on a small dataset, further demonstrating the promise of our approach.

In summary, our work contributes: (1) a novel taxonomy of sound patterns inspired by Deaf-centric ways of describing sounds, and (2) empirical insights from interviews with eight ASL interpreters and two preliminary evaluations showing the initial promise of our pattern-based sound classification approach for enhancing sound awareness of DHH people.

## 2 RELATED WORK

We present background on and situate our work within DHH culture and American Sign Language, sound taxonomies, and current sound awareness solutions.

### 2.1 DHH Culture and American Sign Language

DHH culture encompasses both Deaf and hard-of-hearing communities and possesses a diverse and intricate social fabric that has been explored across numerous academic and cultural disciplines. Deafness is not just audiologically

represented. Researchers have represented hearing loss through three models of disability: the medical, social, and cultural models [5,27,29,38]. While medical and social models emphasize physiological, social, and environmental barriers, the cultural model of deafness embodies a linguistic and cultural group. Individuals in this group belong to the Deaf culture, a diverse cultural milieu characterized by an established set of values, norms, behaviors, and languages like American Sign Language [5,28].

American Sign Language is a natural language with linguistic components like syntax and grammar and is capable of expressing complex and abstract ideas, emotions, and narratives just like another spoken language, but in a visual-spatial modality [21,36]. Pioneered by Stokoe [33], ASL parameters include handshapes, movement, location, palm orientation, and non-manual markers [36]. Non-manual marking (NMM) refers to facial expressions, mouth morphemes, and other body movements that are not made with hands. They provide important grammatical and affective information. For example, brow-lowering indicates WH- questions like "what" and "when." [37]. Another important concept of ASL that is relevant to our current study is classifiers, a morphological system that can represent events and states [12]. Classifiers can be used to represent the entity or describe the size and shape of the objects; therefore, they are helpful in depicting a mental image for the receiver. The construction of classifiers can involve one or both hands. When using both hands, the classifiers often indicate the interactions between objects [12]. This attribute offers flexibility to describe sound events, as sounds can be interpreted as the interactions of materials in an environment [7]. Based on this connection, our study expands the influence of ASL to the development of assistive technologies, beyond supporting communication by exploring how everyday sounds can be effectively classified and described based on sign language interpretations.

### 2.2 Sound Taxonomies and Sound Perception

For decades, researchers have explored systematic categorization schemes for everyday sounds. While various classification frameworks have emerged, the most common one involves differentiating the attributes and semantics of sound sources. Early researchers like Schafer [32] pioneered soundscape research and categorized environmental sounds based on the presence of human activities (*e.g.,* mechanical vs. human sounds). Following this work, many studies innovated domain-specific taxonomies based on the semantics or contexts ("what") of the sound sources, including for urban areas [4,31], restaurants [22], and geographical locations [15].

On the other hand, Gaver's influential research led to an "ecological" categorization approach that is based on the materials of sound sources and their physical interactions with the environment [7]. For example, the sound of a waterfall was described as a *large amount of liquid pouring into a pond from high elevation + high-force splash.* This framework was implemented by Guyot *et al.* [9] through a free sorting task with 30 participants, leading to a taxonomy that classified liquid sounds based on the "discrete" (*e.g.,* drop) and "continuous" interactions (*e.g.,* flow, trickle) with the environment.

Another framework for classifying sounds is based on acoustic signals or audio features, primarily intended for sound classification using native algorithms. For example, Mitrović *et al.* [25] constructed a taxonomy based on the physical and perceptual properties of the sounds in different domains, such as amplitude of the temporal domain and pitch of the frequency domain.

Finally, many recent sound taxonomies employ a hybrid approach that classifies sounds based on both semantic and signal properties [1,8]. Nakatani and Okuno *et al.* [26] developed a taxonomy based on sound sources and sound attributes like timbre and rhythm. Similarly, AudioSet, a 632-class taxonomy, categorizes sounds based on both high-level, semantic relations of sound sources (*e.g.,* animals → pets → dogs) and low-level sound mechanisms, including "sourceless" sound patterns like "whir." This taxonomy aims to cover "all acoustic distinctions" by an everyday listener

and has assisted with the development of multiple sound classification models [10,20]. As another example, Bones *et al.* [1] developed a multi-level taxonomy based on the sound source, subjective perceptions, and explicit acoustic signals. Their study showed that, despite the sound source being the primary cue for categorizing sounds, perception of the acoustic patterns should also be used in order to differentiate multiple sounds originating sources like engines. Notably, in terms of virtual environments, Jain *et al.* [16] interviewed 10 VR sound designers and developed a sound taxonomy along the dimensions of the sound source (*e.g.,* localized speech, animate objects, interaction sounds) and intent (*e.g.,* conveying critical information and aesthetics).

Recent work under the above-mentioned frameworks (*i.e.,* semantics-based, signal-based, and hybrid) has increasingly focused on incorporating human inputs when forming taxonomies [1,8,22]. For example, Bones *et al.* [1] conducted multiple sorting tasks with participants with mixed experience in audio and developed a taxonomy for environmental sounds.

Regardless of the approaches (*e.g.,* source-based and signal-based), most current sound taxonomies are based on the auditory perception and cognition of hearing people. An exception includes Rosen [30], who probed the representations of sounds in American Deaf literature, concluding with two representations: alter-acoustic and extra-acoustic. Alter-acoustic representations (*e.g.,* confusing, frozen) describe the sound experience without "hearist" elements, while extra-acoustic points to the "non-acoustic perceptions" of sounds like touch and rhythm. While these elements inspire our taxonomy, to our knowledge, no research has explored developing sound taxonomies from Deaf-centric perspectives. While developing sound taxonomy for DHH people seems counterintuitive, many prior studies demonstrated the benefits of sound awareness for DHH people to enhance environmental awareness and to help perform daily tasks [2,6,14,17]. Like AudioSet, we hope this novel, Deaf-centric taxonomy can inspire scalable computational approaches for sound awareness.

### 2.3 Sound Awareness Solutions for DHH Users

Early sound awareness solutions for DHH people [13,23,24,35] showed basic visualizations such as spectrographs and "positional ripples" to represent sound properties (*e.g.,* loudness and pitch) and locations. However, they did not have the capability to recognize or distinguish individual sound events – a feature highly desired by DHH people in past studies [2,6]. More recent work started addressing this gap by studying sound recognition systems. For example, Bragg *et al.* designed a preliminary mobile app for sound recognition, but it only covered two sound classes (*i.e.,* alarm and knock), and DHH people desired wider coverage of sounds. Jain *et al.* [17,18] extended this work by developing and deploying at-home and mobile sound recognition systems powered by transfer learning-based deep-CNN models that classify 20 sounds by the sources (*e.g.,* microwave, fire alarms).

However, field studies of these technologies [14,17] showed that the source-based approach is not accurate or reliable enough for everyday use due to numerous factors that may cause misclassification. First, a similar sound source can produce different sounds (*e.g.,* microwave whir when heating the food vs. beeps when it is done). Second, different sound sources can produce similar sounds (*e.g.,* foot stomping vs. book dropping onto the floor). Third, these technologies did not possess contextual information, causing the systems to provide sound feedback that is not appropriate for DHH users' situated contexts. Importantly, misclassifications of sounds due to these factors can be life-critical in certain contexts (*e.g.,* medical devices beeping recognized as appliances [14]). These limitations uncovered the need for involving inputs from DHH users to support sound awareness. These inputs include DHH users' awareness of their surroundings, which is important for recognizing sound events that are appropriate for the contexts.

In this work, we propose a novel sound classification approach that classifies sounds based on ASL's description of sound patterns and presents an 18-class taxonomy of sound patterns that supports this approach. We performed a preliminary evaluation of a state-of-the-art CNN model trained on this taxonomy and found significant improvement over prior models that were trained on source-based taxonomies. Another preliminary evaluation of this approach provided initial evidence that our pattern-based approach can effectively support DHH participants' sound recognition process across different contexts.

## 3 DESIGNING A PATTERN-BASED, DEAF-CENTRIC SOUND TAXONOMY

As we described in Section 2.3, field studies of state-of-the-art sound awareness solutions like SoundWatch uncovered that the real-life accuracy of source-based sound classification approach is prone to many factors, including similar sounding sound sources and evolution of user environments (*e.g.*, old vs. new appliances) [14]. To address this limitation, we propose a pattern-based sound awareness approach that (1) informs DHH users of the underlying patterns of the sound and (2) allows them to infer sound events by incorporating the indicated sound pattern and their knowledge about their situated contexts (*e.g.*, receiving a "liquid running" sound feedback while standing in the kitchen may indicate that the user forget to shut off the water faucet). To implement this approach, we articulated a sound taxonomy based on the underlying patterns of sounds (instead of sound sources). We began by selecting "source-less" sound patterns that represent DHH people's desired sounds (*e.g.*, whir, beep). To ensure that these auditory patterns can be understood by DH people (especially those who rely on visual ways of communicating), we then inquired eight ASL interpreters regarding the signs for the compiled sound patterns and also how they sign sounds to DHH people more generally. Finally, our research team consisting of two ASL interpreters, two CART writers, and four researchers of mixed hearing abilities (1 DHH, 3 hearing) conducted cluster analysis and discussed the responses from eight interpreters and derived the taxonomy that classifies sounds based on how their patterns are signed. We detail the process below.

### 3.1 Method

**Participants:** We recruited 8 ASL interpreters through online study ads, social media (*e.g.*, Reddit posts), emails, and snowball sampling (see Table 1). The average age of these participants was 32.1 years old (*SD*=10.9, *range*=21-50). The average year of experience was 9.25 (*SD*=9.0, *range*=1-29). All interpreters were U.S. residents and had experience working with DHH people professionally.

| PID | Gender | Age | Years of Experience |
|---|---|---|---|
| **I1** | Female | 26 | 5 years |
| **I2** | Female | 50 | 13 years |
| **I3** | Male | 21 | 1 year |
| **I4** | Male | 30 | 8 years |
| **I5** | Female | 23 | 1.5 years |
| **I6** | Female | 26 | 5.5 years |
| **I7** | Male | 34 | 11 years |
| **I8** | Female | 47 | 29 years |

Table 1. Demographics of ASL interpreters for the formative study.

**Procedure:** All interviews were conducted over Zoom videoconference calls by a research team of people with mixed hearing abilities. We proceeded with the sessions once we received the participant's consent with IRB-approved consent forms. At the beginning of the session, we asked participants to complete a brief background form to collect their

demographic information and experience with sign language. We then initiated the protocol containing two parts. For the first part, a semi-structured interview, we asked 15 questions about: (1) interpreters' experience and contexts of working with DHH clients and (2) techniques for interpreting everyday sounds.

For the second part, we wanted to investigate how everyday sounds could be categorized from a Deaf-centric perspective (*i.e.,* ASL) and form categorizations of sounds based on how sign languages convey them. The first and second author, who are hearing, independently selected "source-less" sounds from the Audio Set's sound ontologies [8] that were able to comprehensively represent the sounds in real life. The two authors then met and walked through the two lists. Specifically, the authors identified the common sound items ($N$=13). For sound items that were listed by only one of the two authors, the authors discussed the frequency of events producing these sounds and selected 5 items that were considered "*fairly frequent*" by both authors. During the discussions, we referred to prior findings about desired sounds expressed by DHH people [2,6]. Eventually, we reached a consensus with a list of 18 source-agnostic sounds that were commonly perceived to cover real-life sounds (see supplementary material). We then acquired audio files for these 18 sounds by searching their labels (*e.g.,* "whir") on *FreeSound* [40] and trimmed these audio files so that the maximum length was 5 seconds. After processing the sounds, we uploaded them to *OpenDrive* [41], which allows us to generate direct URLs for the sound clips.

Once we compiled the sounds, we designed a card sorting task (Figure 1) based on *FigJam* [42], a collaborative brainstorming application. We created 18 "sound cards" using *SoundDot* [43], a Figma plugin that allows us to stream sound clips through direct URLs. Each sound card contains a sound clip and a label. To avoid bias, we labeled each sound card with codes (*i.e.,* S1 to S18). We attached the corresponding sound labels in the supplementary material. During the task, we asked participants to click on the sound cards, sign the sounds, and freely move these sound cards and cluster them based on similarities of how these sounds could be signed.

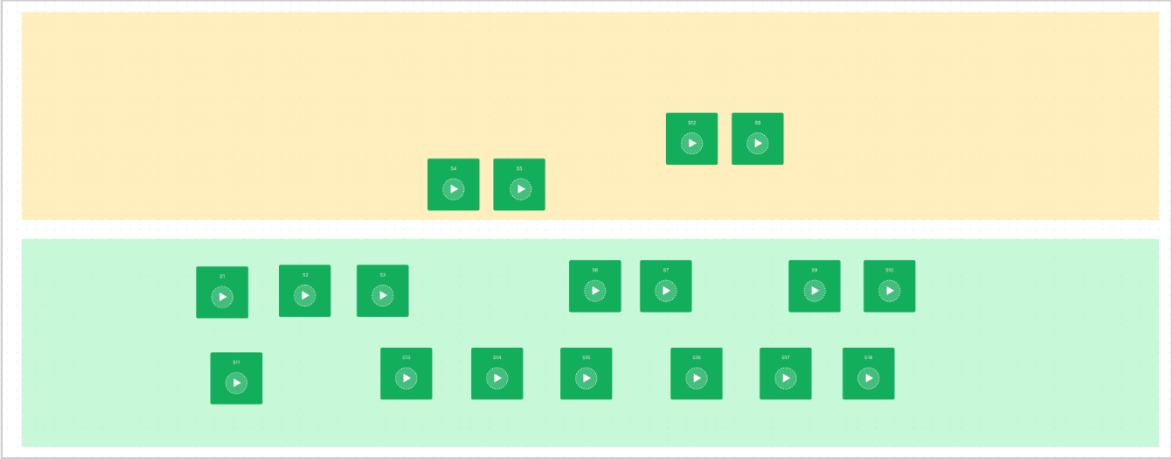

Figure 1. The card sorting task interface. The interface starts with all sound cards in the green section. Participants were asked to cluster sounds that can be similarly signed in the yellow section.

**Data analysis:** Our Study 1 data consisted of the transcripts of the eight interview sessions obtained from real-time captioners and eight copies of FigJam files containing the sorted sound cards. In terms of the transcripts, we used Braun and Clarke's six-phase approach [3]. The first author skimmed and familiarized with the data (step 1) and discussed with the research team to generate an initial codebook (step 2). The first author then walked through the data in detail and

iteratively applied the codes to the data while refining the codebook. The final codebook had a 3-level hierarchy: 6 first-level, 17 second-level, and 63 second-level codes (step 3). The second author independently applied the codes based on the final codebook (step 4). We calculated the interrater reliability between two coders using the ReCal2 package [39] and resolved the disagreements among coders. The average Krippendorff's alpha value was 0.696, and the raw agreement was 84.3%. Finally, we organized the first-level themes (step 5) and constructed our narratives accordingly (step 6). We have attached our final codebook as supplementary material.

| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S1** | | 0 | 1 | 1 | 0 | 4 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | 0 | 1 | 1 | 0 |
| **S2** | 0 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| **S3** | 1 | 1 | | 1 | 0 | 1 | 1 | 0 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 0 |
| **S4** | 1 | 0 | 1 | | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 0 | 1 |
| **S5** | 0 | 0 | 0 | 3 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 1 | 1 | 0 | 1 |
| **S6** | 4 | 0 | 1 | 1 | 1 | | 2 | 0 | 1 | 1 | 5 | 1 | 0 | 2 | 0 | 1 | 1 | 0 |
| **S7** | 1 | 0 | 1 | 0 | 0 | 2 | | 2 | 1 | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| **S8** | 1 | 0 | 0 | 0 | 0 | 0 | 2 | | 0 | 2 | 0 | 3 | 1 | 1 | 0 | 0 | 0 | 1 |
| **S9** | 1 | 0 | 3 | 0 | 0 | 1 | 1 | 0 | | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **S10** | 1 | 0 | 2 | 0 | 0 | 1 | 4 | 2 | 3 | | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| **S11** | 3 | 0 | 1 | 0 | 1 | 5 | 1 | 0 | 1 | 1 | | 2 | 0 | 2 | 0 | 0 | 1 | 0 |
| **S12** | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 0 | 1 | 2 | | 1 | 1 | 0 | 0 | 0 | 1 |
| **S13** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | | 1 | 1 | 0 | 0 | 1 |
| **S14** | 2 | 0 | 1 | 0 | 3 | 2 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | | 1 | 0 | 0 | 0 |
| **S15** | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | | 0 | 0 | 0 |
| **S16** | 1 | 1 | 2 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 2 |
| **S17** | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | | 0 |
| **S18** | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | |

Figure 2. The similarity matrix based on eight participants' sorting task results.

To construct the taxonomy, we performed a cluster analysis on the participants' responses in the card sorting task, which consisted of eight edited FigJam files. We first walked through the individual files and logged the clusters formed by participants. For example, if a participant formed a cluster that contained S7, S10, and S12, we noted "<S7, S10, S12">. After all clusters were logged, we listed all two-item pairs within individual clusters. Using the above example, the three two-item pairs would be:

<S7, S10>, <S7, S12>, <S10, S12>

We then made an 18x18 similarity matrix to visualize the co-occurrences of two-item pairs across eight participants' responses on a Microsoft Excel spreadsheet. For example, if four participants grouped S7 and S12 together, the corresponding cell's value will be 4. We color-labeled the matrix cells based on the cell values to highlight the high-frequency co-occurrences. The similarity matrix appears in Figure 2 above. We considered sound pairs with equal to or higher than 50% agreements (4 interpreters) "signed similarly," which included tap and knock (S6 and S11), beep and tap (S1 and S6), breaking and splash (S7 and S10), and whir and rolling (S5 and S15). The first and last authors then discussed with the two ASL interpreters in the research team to validate these similarities. Per the interpreters' recommendations, we merged the *knock* and *tap* sound into one sound class and did the same for *whir* and *rolling*, leading to a list of 16 sounds with distinct signing patterns. We then asked the two interpreters to verbally describe "*how they would sign these sounds.*" Further discussions with the interpreters elicited two sounds that are common in real life and have distinct signs – "ding" and "whoosh" – and expanded the list to 18 sound classes. We also consulted with two CART writers to assign final class labels by asking them to listen to the audio clips and caption the sounds (*e.g.*, [LIQUID FLOWING]).

### 3.2 Findings

We present the findings based on the analysis of the interview transcripts and sound categorizations from eight ASL interpreters. These findings informed our novel Deaf-centric sound taxonomy, which we also share below.

All participants (*N*=8) stated that they had interpreted everyday sounds when working with Deaf people, with three participants (I2, I4, I8) indicating a high frequency of reporting sounds and that environmental sounds often carried important information. The frequency of interpreting sounds also depended on the context (*N*=5), with three participants stating that interpreting the sounds was especially important for educational (*N*=3), medical (*N*=2), and engineering settings (*N*=1). For example, I5 reflected on her experience working in educational and medical settings:

> "The sounds in the environment like footsteps get interpreted a lot, especially in medical settings because oftentimes there is really important information about who is coming and what is about to happen. When I was doing K-12, there were a lot of environmental sounds that are good information for them to know."

Interpreters reported a diverse set of techniques for describing the sounds in the environments when working with Deaf people, such as explaining the sources and explicitly "*tell what the sound is*" (*N*=4), describing sound characteristics or patterns (*N*=6), and listing possible sounds (*N*=3). Participants demonstrated that sound patterns (*e.g.,* intensity, pitch, and interactions between objects that produced the sound) can be effectively represented with the construction of classifiers (handshapes that represent categories of objects) and non-manual markers (NMM, *e.g.,* facial expressions) (*N*=6). For example, I7 demonstrated that "*rumble*" sounds like "*jet flying by overhead*" can be signed with classifiers and NMM like "widening eyes" and the subtle "blowing air" expression (see Figure 3A). Similarly, for describing "*screechy or squeaky*" sounds, I1 made clear that there are no signs dedicated to the word "*squeak*" but would use both hands with the "CL-B" handshape classifiers, indicate a motion of "*bottom of an object rubbing against a surface,*" and pair it with an NMM of "*harsh sounds that are not pleasant to listen to*" (Figure 3B). The NMMs could represent basic sound characteristics too like the pitch (*N*=6) or could convey the tone and mood of the sounds (I1, I7). For example, I2 and I8 stated that NMM "*puffed cheeks*" may represent "*low-pitched, heavy sounds.*".

Other than the *classifier constructions + NMM* approach, three participants also mentioned that indicating possible familiar sound sources would also help Deaf people make sense of the sound events. For example, when encountering unidentifiable sounds, I6 would *"make a reasonable guess and provide another example or two of what it could be,"* based on the "contexts and what things are happening around." I3 supported this approach, saying that providing possible sound sources could provide a *"reference point"* to help DHH people infer actual sound events based on the context. Importantly, participants pointed out that, for either approach, they would indicate the uncertainty first by prefixing the description with *"it sounds like"*.

We asked participants if any of the sound interpretation strategies they mentioned were recurring strategies followed by other ASL interpreters as well. A small number of participants stated that loud, sudden sounds and some common everyday sounds (e.g., siren) had standard signs (*N*=3). However, most participants reflected that their strategies for interpreting sounds were not taught or based on "*established literature*" (*N*=4), pointing to the need for developing a standard taxonomy. For example, I7 stated:

> "I'm interested to see what [taxonomy] you guys come up with because it is not something that's typically discussed. I haven't seen anything in literature about interpreting sounds [...]. I haven't seen this topic come up at all [in my interpreting school]. So, it's a lot of tribal knowledge."

Similarly, I1 stated that, for "*extremely high-pitched sound*", she might sign something like "*muffling ears*", but "*everyone might do it differently.*" "*That is the thing about interpreting [sounds]. It is never the same,*" she added.

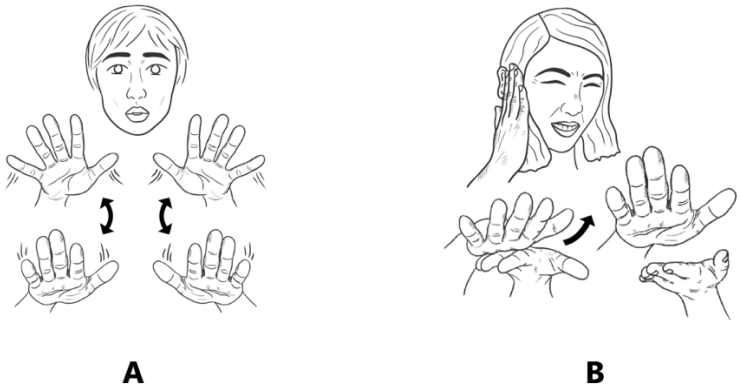


Figure 3. The signs demonstrated by participants who used classifiers + NMM combinations to represent sounds.

Overall, our interviews with eight ASL interpreters reflected on the ways for ASL to indicate sound events. In summary, two approaches emerged:

1. A combination of classifier constructions and non-manual markings (e.g., facial expressions and mouth morphemes) could effectively represent sounds by describing the interactions with the environment and the affective properties of the sound (e.g., pleasant vs. unpleasant, sharp vs. soft).
2. Providing reasonable "reference points" to familiar objects could help DHH people make sense of the sounds based on their environments (e.g., sounds happening in hospital that "*sound like*" *appliance beep* may indicate beeping medical devices).

These insights provided us with guidance to construct a Deaf-centric sound taxonomy inspired by ASL interpretations.

### 3.3 The Deaf-Centric Taxonomy of Sounds

Based on the above insights and following a cluster analysis process described in the "Data analysis" section above, we constructed a taxonomy of sounds. Our Deaf-centric, pattern-based sound taxonomy, outlined in Table 2 below, contains four fields: class code, class labels, ASL descriptions, and examples. Class labels were derived from the captions of sounds noted by two CART writers. ASL descriptions delineated the signing patterns of the corresponding sound classes. The classifiers (*e.g.*, CL-G) or generic objects with dedicated signs (*e.g.*, PAPER) are bolded. ASL descriptions with non-manual markers are also indicated. Based on how interpreters list possible sound events when describing sounds, we include the *examples* field, a column meant for DHH people to get an idea of what sound events they can relate to and form an idea of what the actual sound event may be. For example, knowing that *crumple* sound can be produced by aluminum foils or candy wrappers, a participant may be able to relate to a paper crumpling. We describe more on how DHH people were able to identify specific sound events from our taxonomy in our evaluation below.

| Class | Class Label | ASL Sign Description | Examples |
|---|---|---|---|
| **C1** | Liquid Flowing | **WATER** running down from top | Water coming out of faucet, river flowing, sizzling oil |
| **C2** | Machine Humming | **MACHINE** running<br>(NMM: puffed cheek) | Car engine, dryer |
| **C3** | Shatter | Object **(CL-S)** fall and break suddenly into pieces | Glass bottle breaks, ceramic breaks apart |
| **C4** | Fracture | **STICK (CL-G for both hands)** breaks apart | Breaking branches, chopping wood |
| **C5** | Rip/Tear | **PAPER** being teared apart | Ripping clothes, peeling tape |
| **C6** | Splash | Object **(CL-S)** strike or fall into liquids | Jumping into water, Stepping into mud |
| **C7** | Screech | Indicating a harsh sound from the contact of two surfaces **(CL-B)** | Sudden brake (cars, bicycles), nail on a chalkboard |
| **C8** | Blender | **BLENDER** | Juicer, blender, coffee grinder |
| **C9** | Electrical Buzzer | **BUZZER** | Neon lights, basketball court buzzer |
| **C10** | Beep | **BEEP** | Car horn, microwave beep, fire alarm |
| **C11** | Knock/Tap | Knocking **(CL-S)** on a surface | Door knock, raindrops hitting on the window |
| **C12** | Thump | Dull object **(CL-5)** falls on and hit the surface **(CL-5)** | Footsteps, book falling onto the ground |
| **C13** | Bam/Bang | Sound and vibration of a hard blow | Gunshot, fireworks, thunder |
| **C14** | Scrape | Object **(CL-C)** Scrape or scratch on a surface **(CL-B)** | Chair being dragged, |
| **C15** | Ding/Clink | **BELL** / *downward* CL-A hits CL-B and reverberates | Bell, Toast with wine glasses |
| **C16** | Squeal/Shriek | **SCREAM** | Scream, pig squeals, birds |
| **C17** | Whoosh | Object **(CL-3)** passing with high speed<br>(NMM: thick cheeks blow air out) | Car passing, strong wind |
| **C18** | Crumple | Crush things into wrinkles<br>(NMM: grind teeth) | Aluminum foil, candy wrappers |

Table 2: The Pattern-based, Deaf-Centric Sound Taxonomy.

## 4 PRELIMINARY EVALUATION 1: ONLINE STUDY WITH DHH USERS

The above Deaf-centric sound taxonomy is an integral part of our proposed novel sound classification approach. To evaluate this taxonomy and assess how the descriptions of the sound classes could help DHH people make sense of the sound events in different contexts, we completed a preliminary evaluation activity with nine DHH participants.

### 4.1 Method

**Participants:** Study 2 sessions were conducted by the first, second, and fourth authors, one of whom is DHH. We also recruited an ASL interpreter and a real-time captioner to facilitate communication for all sessions. We proceeded with

the sessions once we received the participants' consent with IRB-approved consent forms. At the beginning of the session, we asked the participant to complete a background form asking about their demographic information, hearing loss, and experience with ASL (see Table 3). The average age of these participants was 38.6 years old (SD=17.6, range=20-72). The average years of experience signing ASL was 24.6 years (SD=18.0, range=2-45).

| PID | Gender | Age | Identity | Hearing loss | ASL Exp. | PMOC |
|---|---|---|---|---|---|---|
| **P1** | Male | 21 | Deaf | Profound | 2 years | Sign Language (D)<br>Sign Language (H) |
| **P2** | Female | 20 | deaf | Moderate | 2 years | Sign Language (D)<br>SimCom (H) |
| **P3** | Female | 33 | deaf | Profound | 32 years | Sign Language (D)<br>Writing (H) |
| **P4** | Female | 30 | Deaf | Moderate | 2 years | Sign Language (D)<br>Sign Language (H) |
| **P5** | Female | 47 | Deaf | Profound | 45 years | Sign Language (D)<br>Writing/Texting (H) |
| **P6** | Male | 72 | Deaf | Severe | 45 years | Sign Language (D)<br>Verbal (H) |
| **P7** | Female | 28 | Deaf | Profound | 26.5 years | Sign Language (D)<br>Writing (H) |
| **P8** | Female | 59 | Deaf | Profound | 30+ years | Sign Language (D)<br>Verbal (H) |
| **P9** | Female | 37 | Deaf | Profound | Whole life | Sign Language (D)<br>Verbal (H) |

Table 3. Study 2 participants' background information. SimCom stands for "simultaneous communication," a communication method where people use spoken language and sign language at the same time. PMOC stands for "preferred mode of communication."

**Procedure:** To assess if our taxonomy can support sound recognition for DHH people, we invited participants to take part in a task where we described the sound patterns and asked participants to infer sound events across different contexts (see Figure 4). The online evaluations were conducted via Zoom conference call. We first acquired 18 sound clips from *FreeSound* [40] by searching with the sound class labels (*e.g.*, "shatter" and "thump"). We avoided referencing sound clips with sound sources to avoid bias. Instead, during the task, we played the sound clip and presented the sound information pertaining to the clip based on our taxonomy: class label, ASL description, and examples. In the future classification system, we intend for the ASL description to be animated in ASL similar to Figure 2 (*e.g.*, on a mobile device or a watch) after recognition. We then asked the participants to list several possible events across different contexts (*e.g.*, kitchen, hiking trails) and provide reasoning for their list. Finally, we asked for any overall thoughts on our taxonomy and our classification approach.

**Data analysis:** Our study 2 data consisted of the transcripts of eight interview sessions and evaluation tasks obtained from the real-time captioner. To analyze interview transcripts, we used the same analysis approach as for Study 1 (Braun and Clarke's six-phase approach as detailed in Study 1 analysis), resulting in a Krippendorf's alpha of 0.692 and raw agreement of 84.8% between the two coders; we attached the final codebook as supplementary materials. For the evaluation tasks, the first and the second authors walked through the transcripts and summarized the participant' inferred sound events for each category of our taxonomy in a table (we detail this table in our findings below).

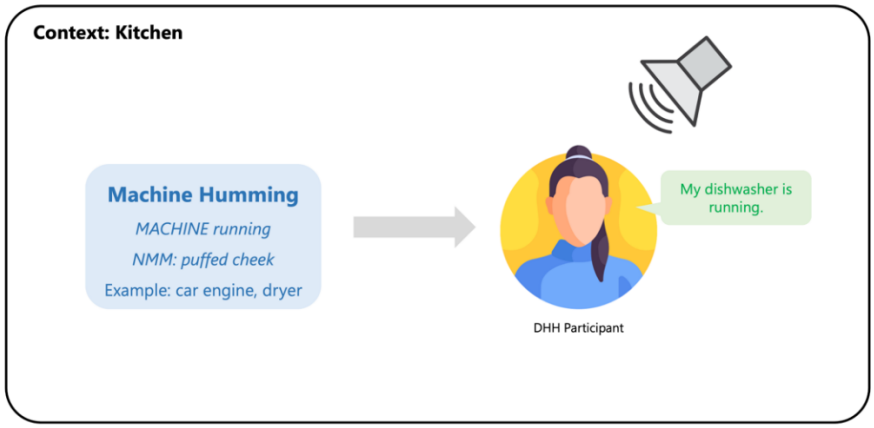


Figure 4. The setup for preliminary evaluations of the novel taxonomy with DHH participants.

### 4.2 Findings

We present the findings from the preliminary evaluation of the pattern-based sound taxonomy with nine DHH signers. Specifically, we investigated whether the ASL descriptions and signs for the sound classes were able to support sound awareness for DHH people across different contexts.

In general, participants appreciated the taxonomy and thought it covered "*most, if not all of*" the everyday sounds they cared about (*N*=8). Participants understood that the source-based sound information might not always be available and indicated that the cues presented by the taxonomy could help them "*know exactly what is happening*" (P2). For example, P5 stated:

> "If I see the '*buzzer*' while [driving] on the road, I would think that there was something wrong with the car."

Similarly, P9 explained that if she was situated in an "unsafe area" and received sound feedback about C13 (*bam/bang*) followed by C16 (*squeal/shriek*), she would infer that as a gunshot and "someone crying."

For our taxonomy evaluation, recall that we presented the participants with three kinds of sound information: audio, ASL descriptions, and example reference events. We asked participants to imagine themselves sitting in various contexts (*e.g.,* kitchen and hiking trail) and infer sound events based on the sound information provided. We summarized the participants' inferred sound events in Table 3 below. For example, we presented C2 (*i.e., machine humming* class) to P6 and asked him: "What sound event do you think it may be if you are at home?" P6 stated that he would think that there is "*a lawn mower running outside.*" Then, for the same C2 class, we asked him to infer possible sound events if he "*saw this sound pattern when standing outside of a farmhouse.*" P6 explained that he would think of "*some kind of motor running,*" therefore inferred "*car engine running*" and "*cotton mills.*"

| Class No. | Class Label | Contexts / Sound Events |
|---|---|---|
| **C1** | Liquid Flowing/Running | Water coming out of faucet (P6), water running (P4, P7, P8, P9), waterfall (P7), rain (P9), river flowing (P9) |

| | | |
|---|---|---|
| **C2** | Machine Humming | Car engine running (P5, P6, P7, P8), washer (P4, P5, P8, P9), dryer (P5), lawn mower (P6, P7), fan (P6), cotton mill (P6) |
| **C3** | Shatter | Fragile object hits the floor (P5), glass bottle breaking (P6, P7, P8), dish breaks (P6), window breaking (P7) |
| **C4** | Fracture | Plastic rod breaks (P5), Wood chopped (P6), Twig snaps (P7), car broken (P9) |
| **C5** | Rip/Tear | Paper being ripped (P4, P6, P7, P8, P9), pants ripping (P4), backpack unzipping (P4, P8), opening a package (P7) |
| **C6** | Splash | Water spilled (P4, P6, P8), waterfall (P6), person jumping into water (P7, P8), heavy rain (P7), drop things into water (P9) |
| **C7** | Squeak/Screech | Whistle (P4), bird calling (P6), feedback from microphone (P7), writing on chalkboard (P9), car screeching to a halt (P9), plane taking off (P9) |
| **C8** | Blender | Blender (P4, P5, P6, P7, P8, P9), coffee grinder (P4, P5, P6), garbage disposal (P5), vacuum cleaner (P6) |
| **C9** | Electrical Buzzer | Alarm (P4, P5, P7), electric stove (P4), Razor (P5), dryer done (P6), phone buzzing (P9) |
| **C10** | Beep | Car honk (P4, P5, P7, P8, P9), microwave done (P4, P7), EKG (P5), phone beep (P5), oven beep (P8), pings for pick-up orders in boba shop (P9) |
| **C11** | Knock/Tap | Door knock (P4, P5, P6, P7, P8), knocking on the countertop (P4), footsteps (P5), settling things down on table (P8), object falls on the pavement (P9) |
| **C12** | Thump | Ball bounce (P4), footsteps (P4, P6, P7, P8), book falling on the floor (P5, P7), tree falls (P8) |
| **C13** | Bam/Bang | Gunshot (P4, P6, P8), door slam (P4), bomb (P5), car accident (P5, P7), fireworks (P7), throwing hammer at wood (P9) |
| **C14** | Scrape/Scratch | Chalk on a chalkboard (P5), scratching on table (P5), moving furniture (P6), scratching on a furniture (P4, P5, P7), scrape on a wok (P8), scratching one's face (P9) |
| **C15** | Ding/Clink | Timer done (P4), wind chimes (P4), toasting with wine glasses (P4), bell (P5, P6, P7), messaging notification (P8), cash register (P8), microwave done (P9) |
| **C16** | Squeal/Shriek | Human scream (P4, P6), animals (P4, P8), baby crying (P5, P8), mouse (P5), bird calls (P6), computer not working (P9) |
| **C17** | Whoosh | Car passing (P4, P5, P6, P7, P8), wind blowing (P5, P9), airplane flying overhead (P7), person walking by (P8), motorcycle passing (P9) |
| **C18** | Crumple | Plastic bags crumpling (P4), paper crumpling (P5, P8), candy wrapper (P6), walking on piles of leaves (P7), foil (P8), crush food into pieces (P9) |

Table 4. A summary of participants' inferred sound events for 18 sound classes.

Participants' responses uncovered several important insights that demonstrated the flexibility and robustness of our taxonomy in supporting sound awareness for DHH people. First, participants' responses showed that our taxonomy could help recognize similar sound events, even if these sound events might possess different sounds in different contexts. For example, while P4 and P7 recognized "microwave done" from the sound class *beep*, P9 inferred this event from *ding*. This convergence makes sense, as the participants' experience with home appliances might be different. For example, old microwaves might have the "ding" sound, but newer, digital ones might produce the "beep" sound instead. This kind of adaptability might be hard to achieve for source-based models. P8 reiterated:

> "Things now can sound very different from 50 years ago. I grew up to hear a window going down. I could feel it 'rolling down,' but now, you hit a button and it has a different sound. These kinds of sounds evolved over time, and through this, I would become more familiar with these sounds."

This finding reflects an insight from a prior evaluation of a source-based sound recognition system in DHH people's homes [17], where a participant reported that remodeling their kitchen made the water running sound undetectable by the system since *"earlier the water hit the porcelain sink, but now I have a stainless-steel sink, and the sound of water hitting is [...] different,"* thereby reinforcing the need for our pattern-based classification approach.

Second, our pattern-based taxonomy helped participants distinguish different sound events, even if these events were produced by the same source. For example, while the door knock sound was almost unanimously categorized into the *knock/tap* class, P4 inferred "door slam" from the *bam/bang* class. Third, the range of participants' inferred sound events (see Table 4) was significantly broader than the number of supported sounds for prior sound recognition systems (about 20 events) [6,17], suggesting that sound recognition pipelines trained on this taxonomy can potentially help DHH users achieve broader sound awareness if they possess knowledge about the situated contexts. Finally, the participants were able to infer sound events that indicate "something went wrong." For example, P9 inferred that *"the car engine could be broken"* from the *fracture* class, while the *machine running* class from prior classification systems helped participants only identify "car engine running" or "car running." Moreover, P7 inferred "microphone feedback" (a loud, often sharp sound that occurs when a microphone picks up and amplifies a sound and feeds it back to a speaker, creating a loop) when presented with information of the *squeak/screech* class.

Even though the feedback was generally positive, a few participants ($N$=3) expressed their concerns. For example, P9 pointed out that this taxonomy lacked classes that represent siren sounds like "*those for hurricanes.*" Considering that sirens had distinct sound patterns that separated themselves from others, we propose the addition of another sound class, *siren*, that represents emergency sirens like fire trucks and police cars. P6 suggested that "*there [might be] a certain amount of training,*" indicating the potential learning curve of this taxonomy. However, P6 expected this learning process to be short, and P4 stated that using ASL to describe the sound patterns was "extremely helpful" for understanding the sound events. This reiterates our belief that our taxonomy is just the beginning of the exploration of this Deaf-centric, pattern-based sound awareness approach.

## 5 PRELIMINARY EVALUATION 2: SOUND CLASSIFICATION EXPERIMENT

While the preliminary evaluation studies with DHH people showed initial promise in the taxonomy of sound patterns, our classification approach is only effective if the patterns can be classified and distinguished algorithmically. Thus, in the second evaluation, we trained and evaluated a sound classification model based on our taxonomy.

To create our model, we first compiled our dataset of sound patterns. To do so, we downloaded sound clips for the 18 categories in our taxonomy from *FreeSound* [40], an online corpus of high-quality, labeled sound effects. All downloaded clips were converted to a single format (16Khz, 16-bit, mono), and silences greater than one second were removed, resulting in 9.8 hours of recordings. We divided our recordings into a train and a test set, with 80% and 20% split, respectively.

To generate input features for our model, we segmented each clip into one-second segments and computed short-time Fourier Transforms using a 25ms sliding window and 10ms step size (frequency range 20Hz to 8000Hz), yielding a 96-length spectrogram. We then converted our linear spectrogram into a 64-bin log-scaled Mel spectrogram and

generated a 100 X 64 input frame for every second of audio. To these log-mel spectrograms, we applied Cepstral Mean and Variance Normalization (CMVN) [34].

To train our model, we adopted a transfer learning approach commonly used for sound classification (*e.g.,* [17,18,20]). We downloaded a pre-trained *VGG-16* CNN model [11], replaced the last fully connected layer with a fresh layer (using a sigmoid activation function), and fine-tuned the model on our training set. For training, we used a cross-entropy loss function with an Adam optimizer [19].

To evaluate our model, we used a clip-level prediction. More specifically, we aggregated the classification confidences for each one-second prediction across the entire clip and returned the top prediction. We found that our model returned a near-perfect accuracy of 98.6% on our test set, showing the promise of our pattern-based taxonomy for accurately distinguishing sounds.

However, our evaluation was quantitative and performed on a small dataset. For a more accurate assessment, future studies should conduct evaluations over bigger datasets and in the field with DHH users in a variety of contexts.

## 6 DISCUSSION

Inspired by the limitations of the source-based sound awareness solutions, we investigated a novel sound awareness approach that builds on DHH users' contextual awareness and informs them of pattern-based sound information to enhance their sound awareness. As the first step for implementing this approach, we compiled a list of sound patterns that encapsulate all sound events desired by DHH people and asked eight ASL interpreters to sign, sort, and categorize them based on how these sound patterns were signed in ASL. Our multidisciplinary team, consisting of 1 DHH researcher, 3 hearing researchers, 2 CART writers, and 2 ASL interpreters, conducted cluster analysis on the interpreters' responses and collaboratively articulated a novel taxonomy that classifies sound patterns based on their sign descriptions. This 18-class taxonomy contains four fields: class code (*e.g.,* C3), class label (*e.g., fracture)*, ASL descriptions (*i.e.,* verbal description of how this sound pattern can be signed), and example 'reference' sound events.

To our knowledge, this is the first 'Deaf-centric' taxonomy that can be used for classification purposes and sound awareness support for DHH people. While our taxonomy is distinct from the prior source-based ontologies (*e.g.,* [22,31,32]), we argue that our "pattern-based" classification approach extends the "hybrid approach" category we introduced in our Related Work section 2.2 since it compares with the "ecological" [7,9], signal-based approaches [25], and Bone's "middle-level" and "low-level" sound taxonomy [1] in two ways. First, the ASL descriptions of many sound classes in our taxonomy delineate the interactions of generic objects and materials (*e.g.,* paper) that produced the sound, an attribute that characterizes Gaver's taxonomy [7]. Second, some class labels and ASL descriptions reflect explicit judgments of the sound characteristics (*e.g.,* continuous vs. discrete), a feature used by Mitrović *et al.* [25] and Bones *et al.* [1] to classify mechanical sounds. Third, sound classes that include NMM as part of the ASL descriptions contain affective properties of the sound (*e.g.,* harsh, unpleasant), which resembles the "subjective state" feature used by Bones *et al.* [1] to classify dog sounds (*e.g.,* "aggressive" and "deep").

To evaluate this taxonomy and its ability to support DHH people's sound awareness, we conducted two preliminary evaluations to answer two questions:

3. Is sound information from the taxonomy able to support DHH participants' sound inference process across different contexts?
4. Can the sound-pattern classes in our taxonomy be recognized and distinguished algorithmically?

In terms of the first question, participants' responses provided initial evidence that DHH people were able to recognize sound events across different contexts based on our taxonomy's ASL descriptions. Compared to source-based approaches widely used by prior sound awareness solutions (*e.g.,* [17,18,20]), our pattern-based taxonomy offers high flexibility by enabling DHH people to recognize different sound events in different contexts. Moreover, the pattern-based sound classification approach can adapt to the evolution of user environments (*e.g.,* old vs. new microwave indicator sounds), while this adaptiveness can be hard to achieve with source-based approaches. To answer the second question, we trained and evaluated a sound classification model based on the taxonomy on a small dataset. The model evaluation demonstrated near-perfect accuracy (98.6%), a notable improvement from the state-of-the-art (*e.g.,* SoundWatch [18]), showing the promise of this pattern-based sound classification approach.

Our work opens up new possibilities for future human-AI systems to support sound awareness for DHH people. Instead of identifying individual sound sources, whose accuracy is prone to be negatively influenced by factors like similar sounding sounds and changes in contexts [14], future sound awareness systems can utilize the strength of human (contextual awareness) and machine intelligence (pattern recognition) to help DHH people be aware of their surroundings and perform everyday tasks. Moreover, beyond the scope of our findings, we believe that our approach offers DHH people more end-user agency to teach, customize, and personalize the sound awareness systems (*e.g.,* assigning labels to recognized patterns), which, according to prior studies, can be empowering for DHH people [14].

We imagine our work to be applied in many domains. First, the novel 18-class taxonomy is a step toward establishing standard vocabularies for describing everyday sounds with sign language, as many interpreters have expressed interest in its potential to contribute to ASL. Second, the pattern-based sound classification approach can be used to help DHH people perform tasks in both everyday and professional settings. For example, future sound awareness tools based on the pattern-based sound classification models can help DHH users detect anomalies with their cars by identifying sound patterns in real-time (*e.g.,* sound class changing from *machine running* to *screech/squeak* while driving might indicate an issue with brake disc). Similarly, in professional engineering settings, a real-time pattern-based sound classification system can help mechanics who are DHH perform inspection by keeping them updated on current sounds produced by the machines.

Future work should investigate the contributions of our pattern-based sound classification approach towards the vision of a complete, semantic awareness of the sounds for DHH people by developing sound awareness technologies (*e.g.,* mobile apps) that enable DHH users to interact with the pattern-based sound feedback. For example, can DHH users use this technology to keep track of the working status of the washers and dryers? Can this approach help DHH people inspect if the air conditioning unit is running? Many questions remained about the real-life efficacy of our approach in supporting broader and nuanced awareness of DHH users' surroundings.

### 6.1 Limitations and Future Work

Our study has several limitations. First, all eight interpreters taking part in our interviews specialized in ASL. It is important to note that sign language is not universal, and all regional sign languages may have different lexicons, grammar, and cultural nuances. Besides the ASL-specific components like classifiers, our taxonomy is generalizable, but future work should evaluate the taxonomy with interpreters specializing in other languages (*e.g.,* Indo-Pakistani Sign Language, British Sign Language).

Second, both of our evaluations are preliminary and can only provide initial evidence for our claims. In terms of the online evaluation with nine DHH participants, we did not involve any tangible prototypes applying the pattern-based sound classification approach. Similarly, the quantitative sound classification experiment, where we trained a model

based on the sound classes presented in the taxonomy, only involved a small dataset. Moreover, this model was not deployed in the field. Future work should conduct this experiment with a large-scale dataset and use a deployable, end user-friendly application to test the real-world accuracy of the model.

Third, the DHH population consists of different cultural sub-groups, and not all may know sign language. Fortunately, our taxonomy describes sounds in multiple formats (class names based on audible patterns, ASL description, and example reference sound events), which, while crucially inspired by ASL, cover a broad spectrum of auditory and non-auditory ways of perceiving sounds. Still, future work should evaluate our taxonomy with deaf, hard-of-hearing, and Deaf groups separately to verify generalizability.

Finally, while our classification approach is heavily informed by DHH perspectives, we acknowledge that not all DHH people may want to know about sounds and that our approach may not work as designed for everyone. However, several past studies [2,6,17], including a large-scale survey with 201 DHH participants [6], support that many DHH people want increased sound awareness. Owning to the flexibility of our taxonomy, future work should investigate personalized sound recognition systems that can be constrained to detect a small subset of patterns from our taxonomy (*e.g.,* those that relate to safety, such as 'bam/bang') while otherwise avoiding the hearing world.

## 7 CONCLUSION

To enhance sound awareness of DHH users, researchers have built sound classification systems to distinguish different sound sources (*e.g.,* microwave, dishwasher) or sound events (*e.g.,* door knocks). However, since multiple objects can produce similar sounds, these systems often classify inaccurately, limiting their utility for DHH people. In this work, we investigate a novel pattern-based, Deaf-centric approach for classifying sounds. We began with interviewing eight ASL interpreters on how they interpret sounds to DHH people and, based on the findings, articulated an 18-category taxonomy of signed sound patterns. We then performed two preliminary evaluations of our taxonomy, finding initial promise for our two classification goals: that different classes of our taxonomy can be recognized accurately and that DHH people are able to obtain their desired sound information from our taxonomy. We outline future work for building robust, scalable, and personalized sound recognition systems based on our classification approach.

## REFERENCES


[1] Oliver Bones, Trevor J. Cox, and William J. Davies. 2018. Sound Categories: Category Formation and Evidence-Based Taxonomies. *Front. Psychol.* 9, (July 2018), 1277. DOI:https://doi.org/10.3389/fpsyg.2018.01277

[2] Danielle Bragg, Nicholas Huynh, and Richard E. Ladner. 2016. A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Reno Nevada USA, 3–13. DOI:https://doi.org/10.1145/2982142.2982171

[3] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide*. SAGE Publications.

[4] A. L. Brown, Jian Kang, and Truls Gjestland. 2011. Towards standardization in soundscape preference assessment. *Appl. Acoust.* 72, 6 (May 2011), 387–392. DOI:https://doi.org/10.1016/j.apacoust.2011.01.001

[5] Anna Cavender and Richard E. Ladner. 2008. Hearing Impairments. In *Web Accessibility: A Foundation for Research*, Simon Harper and Yeliz Yesilada (eds.). Springer, London, 25–35. DOI:https://doi.org/10.1007/978-1-84800-050-6_3

[6] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300276

[7] William W. Gaver. 1993. What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecol. Psychol.* 5, 1 (March 1993), 1–29. DOI:https://doi.org/10.1207/s15326969eco0501_1

[8] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, New Orleans, LA, 776–780. DOI:https://doi.org/10.1109/ICASSP.2017.7952261

[9] Patrice Guyot, Olivier Houix, Nicolas Misdariis, Patrick Susini, Julien Pinquier, and Régine André-Obrecht. 2017. Identification of categories of liquid sounds. *J. Acoust. Soc. Am.* 142, 2 (August 2017), 878–889. DOI:https://doi.org/10.1121/1.4996124

[10] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. 2017. CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 131–135. DOI:https://doi.org/10.1109/ICASSP.2017.7952132

[11] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. 2017. CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 131–135. DOI:https://doi.org/10.1109/ICASSP.2017.7952132

[12] Joseph Hill, Diane Lillo-Martin, and Sandra Wood. 2018. *Sign Languages: Structures and Contexts* (1st edition ed.). Routledge, London ; New York.

[13] F Wai-ling Ho-Ching, Jennifer Mankoff, and James A Landay. Can you see what I hear? The Design and Evaluation of a Peripheral Sound Display for the Deaf.

[14] Jeremy Zhengqi Huang, Hriday Chhabria, and Dhruv Jain. 2023. "Not There Yet'': Feasibility and Challenges of Mobile Sound Recognition to Support Deaf and Hard-of-Hearing People. *N. Y.* (2023).

[15] Lingjiang Huang and Jian Kang. 2015. The sound environment and soundscape preservation in historic city centres—the case study of Lhasa. *Environ. Plan. B Plan. Des.* 42, 4 (July 2015), 652–674. DOI:https://doi.org/10.1068/b130073p

[16] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. A Taxonomy of Sounds in Virtual Reality. In *Designing Interactive Systems Conference 2021*, ACM, Virtual Event USA, 160–170. DOI:https://doi.org/10.1145/3461778.3462106

[17] Dhruv Jain, Kelly Mack, Akli Amrous, Matt Wright, Steven Goodman, Leah Findlater, and Jon E. Froehlich. 2020. HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376758

[18] Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. SoundWatch: Exploring Smartwatch-based Deep Learning Approaches to Support Sound Awareness for Deaf and Hard of Hearing Users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Virtual Event Greece, 1–13. DOI:https://doi.org/10.1145/3373625.3416991

[19] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. Retrieved September 14, 2023 from http://arxiv.org/abs/1412.6980

[20] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, ACM, Berlin Germany, 213–224. DOI:https://doi.org/10.1145/3242587.3242609

[21] Scott K. Liddell. 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press, Cambridge. DOI:https://doi.org/10.1017/CBO9780511615054

[22] PerMagnus Lindborg. 2016. A taxonomy of sound sources in restaurants. *Appl. Acoust.* 110, (September 2016), 297–310. DOI:https://doi.org/10.1016/j.apacoust.2016.03.032

[23] Tara Matthews, Janette Fong, F. Wai-Ling Ho-Ching, and Jennifer Mankoff. 2006. Evaluating non-speech sound visualizations for the deaf. *Behav. Inf. Technol.* 25, 4 (July 2006), 333–351. DOI:https://doi.org/10.1080/01449290600636488

[24] Tara Matthews, Janette Fong, and Jennifer Mankoff. 2005. Visualizing non-speech sounds for the deaf. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, ACM, Baltimore MD USA, 52–59. DOI:https://doi.org/10.1145/1090785.1090797

[25] Dalibor Mitrović, Matthias Zeppelzauer, and Christian Breiteneder. 2010. Chapter 3 - Features for Content-Based Audio Retrieval. In *Advances in Computers*. Elsevier, 71–150. DOI:https://doi.org/10.1016/S0065-2458(10)78003-7

[26] T. Nakatani and HIroshi G. Okuno. 1998. Sound Ontology for Computational Auditory Scence Analysis. Retrieved August 30, 2023 from https://www.semanticscholar.org/paper/Sound-Ontology-for-Computational-Auditory-Scence-Nakatani-Okuno/c81832ddcaba13f595510b8338f40fabf535ebbb

[27] Michael Oliver. 1996. *Understanding Disability*. Macmillan Education UK, London. DOI:https://doi.org/10.1007/978-1-349-24269-6

[28] Carol Padden and Tom Humphries. 1990. *Deaf in America: Voices from a Culture*. Harvard University Press, Cambridge, MA.

[29] D. Power. 2005. Models of Deafness: Cochlear Implants in the Australian Daily Press. *J. Deaf Stud. Deaf Educ.* 10, 4 (July 2005), 451–459. DOI:https://doi.org/10.1093/deafed/eni042

[30] R. S. Rosen. 2007. Representations of Sound in American Deaf Literature. *J. Deaf Stud. Deaf Educ.* 12, 4 (April 2007), 552–565. DOI:https://doi.org/10.1093/deafed/enm010

[31] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. 2014. A Dataset and Taxonomy for Urban Sound Research. In *Proceedings of the 22nd ACM international conference on Multimedia* (MM '14), Association for Computing Machinery, New York, NY, USA, 1041–1044. DOI:https://doi.org/10.1145/2647868.2655045

[32] R. Murray Schafer. 1993. *The Soundscape*. Retrieved September 1, 2023 from https://www.simonandschuster.com/books/The-Soundscape/R-Murray-Schafer/9780892814558

[33] W. C. Stokoe. 2005. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *J. Deaf Stud. Deaf Educ.* 10, 1 (January 2005), 3–37. DOI:https://doi.org/10.1093/deafed/eni001

[34] O. M. Strand and A. Egeberg. 2004. Cepstral mean and variance normalization in the model domain. Retrieved September 14, 2023 from https://www.semanticscholar.org/paper/Cepstral-mean-and-variance-normalization-in-the-Strand-Egeberg/0de27e275803a000babcfa5c06c0683ee1df76e0

[35] M Tomitsch and T Grechenig. DESIGN IMPLICATIONS FOR A UBIQUITOUS AMBIENT SOUND DISPLAY FOR THE DEAF.

[36] Clayton Valli and Ceil Lucas. 2000. *Linguistics of American Sign Language: An Introduction*. Gallaudet University Press.

[37] Katharine L Watson. 2010. WH-questions in American Sign Language: Contributions of non-manual marking to structure and meaning. *Theses Diss. Available ProQuest* (January 2010), 1–130.

[38] A. M. Young. 1999. Hearing parents' adjustment to a deaf child-the impact of a cultural-linguistic model of deafness. *J. Soc. Work Pract.* 13, 2 (November 1999), 157–176. DOI:https://doi.org/10.1080/026505399103386

[39] 2010. ReCal2: Reliability for 2 Coders – Deen Freelon, Ph.D. Retrieved September 11, 2023 from http://dfreelon.org/utils/recalfront/recal2/

[40] Freesound - Freesound. Retrieved September 11, 2023 from https://freesound.org/

[41] Unlimited Cloud storage | Cloud Backups | Cloud Drive. Retrieved September 11, 2023 from https://www.opendrive.com/

[42] The Online Collaborative Whiteboard for Teams. *Figma*. Retrieved September 11, 2023 from https://www.figma.com/figjam/

[43] SoundDot for LabKit | Figma Community. *Figma*. Retrieved September 11, 2023 from https://www.figma.com/community/widget/1094610197891170835/SoundDot-for-LabKit