ROGER: Visualizing Voice Records to Enhance Team Communication Trainings for High-Stress Situations

Michael Oppermann
michael.oppermann@ait.ac.at
AIT Austrian Institute of Technology
Vienna, Austria

Jakob Carl Uhl
AIT Austrian Institute of Technology
Vienna, Austria
Paris Lodron University of Salzburg
Salzburg, Austria

Georg Regal
AIT Austrian Institute of Technology
Vienna, Austria
Paris Lodron University of Salzburg
Salzburg, Austria

Manfred Tscheligi
AIT Austrian Institute of Technology
Vienna, Austria
Paris Lodron University of Salzburg
Salzburg, Austria

Markus Murtinger
AIT Austrian Institute of Technology
Vienna, Austria
Paris Lodron University of Salzburg
Salzburg, Austria

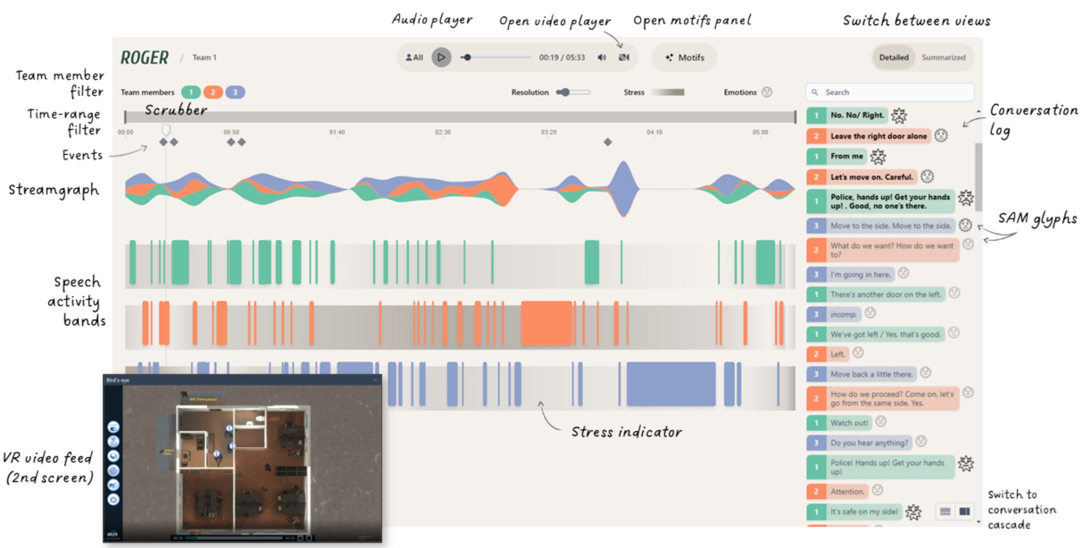

Figure 1: *ROGER*: Interactive visualization and motif-querying interface for analyzing team communication. *Center*: comm. patterns; *Right*: conversation log; *Top*: various controls. Detailed annotations in the screenshot highlight individual components.

Abstract

Effective communication is essential in high-stress environments but stress often disrupts the flow of information and leads to miscommunication. While scenario-based training exercises are widely used, post-hoc reflection and analysis of verbal interactions remain challenging due to overlapping speech, limited analysis time, and the dynamic nature of these situations. This paper introduces *ROGER*, a novel visual analytics interface designed to support after-action reviews of communication during high-stress training scenarios. Developed in collaboration with police trainers through an iterative design study, *ROGER* integrates emotional voice metrics, heart rate variability, and spoken language content to provide a comprehensive analysis of team communication. The system enables a flexible in-depth exploration of communication patterns through *motifs*—repeated sequences or content elements—including those generated by a large language model (LLM) as well as predefined ones. Our approach addresses the limitations of existing tools, which focus primarily on content summarization or voice replays without incorporating emotional and stress-related voice data. We validated the utility through interviews with police trainers and conducted a workshop with medical first responders to investigate the potential for cross-domain applicability. Our findings provide preliminary evidence that *ROGER* supports effective team performance analysis in diverse high-stress environments. See also supplemental material at https://osf.io/pc6un

This work is licensed under a Creative Commons Attribution 4.0 International License.
VINCI 2025, Linz, Austria
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1845-8/25/12
https://doi.org/10.1145/3769534.3769554

CCS Concepts

• **Human-centered computing** → **Visualization systems and tools**; **Visualization techniques**.

Keywords

Visualization, Communication, LLM-assisted analysis

ACM Reference Format:

Michael Oppermann, Jakob Carl Uhl, Georg Regal, Manfred Tscheligi, and Markus Murtinger. 2025. *ROGER*: Visualizing Voice Records to Enhance Team Communication Trainings for High-Stress Situations. In *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (VINCI 2025), December 01–03, 2025, Linz, Austria.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3769534.3769554

1 Introduction

The ability to communicate clearly is essential in high-stakes environments like law enforcement, emergency response, and military [14]. In these settings, team members must convey and interpret essential information quickly and accurately under intense pressure. However, stress significantly hinders the retrieval of learned skills and often leads to miscommunication and errors [10]. Repetition and reflection in team training are essential for internalizing successful communication strategies. In this work, we focus on police tactical training, where a common method is scenario-based training. In these exercises, teams collaboratively resolve realistic, high-stress scenarios that closely mimic real-life operations, aiming to develop and refine a broad set of skills.

Recalling specific interactions can be challenging for both participants and trainers due to the dynamic nature of these situations [32]. Despite the recognized importance of communication, current training tools fall short in providing detailed post-hoc analyses of verbal communication dynamics. The predominant method remains the manual review of recordings, where trainers and trainees replay audio to reconstruct interactions. Beyond very limited tool support, consistent standardized guidelines for communication analysis are lacking. While the use of specific terminology and role assignments is critical, these situations are fluid and require flexible, adaptive communication strategies. Moreover, stress and emotions significantly influence communication performance and should be incorporated into its analysis [21].

Existing visualization approaches in this context have primarily focused on summarizing conversations through keyword extraction and topic modeling. While these methods provide useful overviews of conversations and thematic structures, they do not fully account for the unique challenges of communication training in high-stress conditions, which require a more flexible analysis approach.

To address these challenges, we designed and implemented *ROGER*, a novel visualization system to enhance team communication training in high-stress environments. Developed through an iterative visualization design study [35] in collaboration with police trainers, *ROGER* integrates heart rate variability (HRV) records, emotional voice metrics, spoken content, and contextual data to support a comprehensive analysis of team communication. We introduce a motif-based querying approach for flexible, interactive analysis by identifying communication patterns or content elements—including predefined motifs and custom motifs assisted by a large language model (LLM)—to compensate for the inherently dynamic nature of team communication trainings. We contribute:

- An analysis and abstraction of data and challenges for analyzing voice records in the context of team communication training.
- A conceptual motif model for analyzing the quality of verbal communication.
- The design and implementation of *ROGER*, a visualization and motif-querying interface for analyzing voice metrics, emotional expressions, and spoken language content.

We validated the utility of *ROGER* through engagement with police trainers and conducted a workshop with medical first responders to explore its applicability in other high-stakes environments. These sessions provided preliminary evidence of *ROGER*'s effectiveness in enhancing after-action communication reviews of team trainings and underscore the potential of an intelligent, LLM-supported analysis for verbal communication.

2 Related Work

We review relevant previous work on team communication and the analysis of voice records, followed by a discussion of existing techniques for visualizing conversations, stress, and emotions.

2.1 Analyzing Team Communication

Team communication has been extensively studied due to its crucial role in task coordination and performance [42]. Research has shown that the quality and frequency of communication are directly related to team performance, with high-performing teams demonstrating more effective communication patterns compared to low-performing teams [33]. Stroomer et al. [41] examined how analyzing communication in team tasks provides insights into team dynamics and decision-making processes. Speech patterns have been identified as key indicators of performance and offer meaningful insights into how teams function under pressure [44]. Baber et al. [1] used speech act theory to monitor team communications during training exercises. Recent work has also leveraged natural language processing techniques, including deep learning-based models, to identify dialogue acts and communication behaviors [25, 28]. While previous studies provide valuable insights into team communication and performance, they do not address the need for a dedicated analysis interface designed for after-action reviews of team communication— the primary objective of our work.

2.2 Visualizing Conversations

Visualization techniques to summarize conversations from diverse settings have been well studied in recent years. Several works explored how to compute topics of conversations as visual summaries [7, 36]. The ThemeRiver[12] approach is commonly used to show the temporal evolution of topics. El-Assady et al. [9] presents computational methods and visualizations for analyzing argumentation patterns across topics and time. ConVis [16] uses both topic and opinion mining results to support the interactive exploration and navigation of blog conversations. However, in our usage context, topic-based analysis proves too broad. In most cases, verbal interactions during high-stress training are brief and the topics are predefined. Instead, it is essential to perform a more granular analysis of the spoken words and to incorporate additional contextual data.

Several tools have focused on visualizing conversations within meetings to foster creativity or support balanced communication [22], such as IdeaWall [37], which extracts keywords and augments them with web-search results. In contrast, our focus is on after-action reviews and the detailed analysis of completed training sessions, rather than real-time conversation facilitation.

Matur et al. [23] suggested to incorporate pitch and volume in conversation visualizations and emphasized the importance of vocal characteristics. We build upon those tools, including Conversation Clock [2] and Flow Client [45], to depict the interaction rhythm and conversational patterns. *ROGER* extends these concepts by incorporating additional data, including emotional metrics, and showing transcripts side by side.

2.3 Visualizing Stress and Emotional Characteristics

A holistic analysis of team communication incorporates stress and emotions. Affective data visualizations have been explored in various contexts. Several works have investigated how to visualize collective stress as an intervention mechanism, using methods such as augmented reality [47], circular bar plots [20], and dynamic abstract paintings [40]. Others have focused on personal stress visualizations [31, 38, 39], which aim to provide individual insights into stress levels. Research has also examined the visualization of emotions to analyze and communicate affective states [46]. However, the design space for visualizing stress and emotions in the context of verbal team communication remains under-explored. Previous work has primarily focused on summarization techniques, whereas our approach aims to visualize emotional states on a per-message basis. *ROGER* is the first system to provide a comprehensive post-hoc analysis that directly links spoken content with stress and emotional characteristics.

3 Design Process

Here, we outline *how* we conducted our research while we describe domain challenges in Sec. 4 and reflect upon on our methodological decisions in Sec. 9. *ROGER* was designed and implemented in a highly iterative process, through multiple rounds of engagement with police trainers and experts.

A previous **field study** *[cit. anonymized]*, which explored the use of virtual reality (VR) in police enforcement training, highlighted the need to address communication challenges. Police officers rely on verbal communication for coordination, situational awareness, and de-escalation, yet trainers and trainees struggle to analyze it with existing tools, especially given the limited training time available to practice multiple skills beyond communication.

Building on these insights, we organized an **elicitation workshop** with police trainers and law enforcement experts. For this workshop, part of a police symposium, we recruited eight participants who were not involved in the field study. The workshop followed a semi-structured format, during which we gathered expert statements and organized them into topics. These included discussions on what constitutes effective communication in stress situations, the difficulties in evaluating communication, and the potential role of visualization and technologies in enhancing training.

As part of an iterative process, we conducted further **expert interviews** to refine our data and task abstractions, as well as inform interface decisions and interaction paradigms. We created data sketches using example voice recordings collected during the field study. These sketches and prototypes served as discussion tools to gather feedback from five police trainers and training leaders from Germany, Switzerland, and the Netherlands. Three participants were engaged from our existing network, while two were recruited through snowball sampling. To ensure fresh perspectives, none of these experts had attended the elicitation workshop.

All five experts had served in various roles at the police before transitioning to instructional positions, and thus brought valuable practical insights to the design process. Unfortunately, all participants were male and reflect the broader reality of police training where female trainers remain a minority [8].

To further explore the potential cross-domain applicability of *ROGER*, we organized a 1.5-hour **adaptation workshop** with five experts in medical emergency response. All participants are involved in training programs and evaluating new technologies. The workshop goal was to investigate the broader utility and assess whether there are unique requirements in the medical communication context. We summarize the design implications in Sec. 8.

4 Domain Background and Challenges

Our motivation stems from the need to improve internal police communication, particularly to support trainers and trainees during after-action reviews. Teams in these high-stress training scenarios typically consist of 2-7 members, with 3-member teams being the most common. In specialized crisis simulations, larger groups are involved, although they are generally subdivided into smaller teams. Each training session usually runs for 5-10 minutes. In the initial workshop and interviews, we gathered a large set of situations and patterns that lead to communication failures. Based on this problem elicitation, we identified key challenges in verbal team communication (C1-C4) and its analysis (C5-C7), which we abstracted into domain-agnostic challenges to ensure transferability.

C1: Diverse backgrounds. Teams often consist of participants from different agencies and organizations, with distinct goals, norms, and procedures. Members frequently do not know each other, which affects communication dynamics.

C2: Impact of stress and emotions. Stress and emotions significantly affect the performance. Deciding when and how to communicate becomes challenging, especially in high-stress situations where actions may need to precede communication. Certain trainings are intentionally designed to introduce stressors to practice decision-making but also to enhance team communication under pressure [27, 43].

C3: Impact of events. Communication performance may change drastically after events, such as a shot being fired or an accident occurring. These high-stakes moments, typically linked with C3, can disrupt the flow of communication and dictate the outcome.

C4: Fuzzy guidelines. The complexity and dynamic nature of the situations often prevent the establishment of clear and stable communication guidelines [15]. Different organizations also have varying degrees of strictness in their protocols, and these can change rapidly based on regional or legal contexts. Specific terms and phrases are used in some situations, but high-stress environments make consistent adherence to protocols difficult, unlike structured settings such as emergency hotlines.

C5: Contextual analysis. Communication should not be analyzed in isolation but rather within the context of the entire scenario. The meaning and effectiveness of verbal exchanges are closely linked to events, locations, and other factors.

C6: Recall difficulties. Team members often struggle to recall what they have said during high-stress scenarios. They can usually recount the sequence of events but may not remember specific phrases or formulations accurately, and there may even be disagreements about what was or wasn't communicated. To mitigate this lack of clarity, voice recordings are created more frequently, but without specialized analysis tools, the process of reviewing communication is time-consuming and cumbersome.

C7: Limited analysis time. Communication is just one of many aspects that are practiced simultaneously and need to be analyzed and reflected upon. As a result, trainers and trainees have very limited time for analysis and to derive actionable insights.

Following a deeper understanding of the problem context and challenges, we defined three high-level analysis tasks:

T1: Analyze the spoken language content.
T2: Analyze the delivery of messages.
T3: Analyze the dynamics of team communication.

Each analysis task may incorporate the impact of events and contextual factors, as well as the influence of stress and emotions. Existing approaches do not support this level of comprehensive analysis. *ROGER* fills this gap by addressing these three core analytical tasks.

5 Data

We use voice records, heart rate measures, and contextual data as raw data sources. We describe how we collected and processed the data, followed by a detailed data characterization.

Data was collected during an operational training conducted by a European police authority in 2023, with the primary goal to study the use of VR. The training context was on high-stress conditions and the standardized scenarios were developed by the authority and not the research team. The training used the Refense system [29] for VR-based replays and analysis. A total of 18 teams (63 police officers; 53 male, 10 female) of 3–4 members each completed a 5-10 minute training session. Each trainee was equipped with a microphone (RØDE smartLav+) and a heart-rate chest sensor (Polar H10). In addition to individual VR video replays for each participant, a bird's-eye view of the entire team was recorded. Timestamps were logged when shots were fired to mark critical events.

For further data processing, we randomly selected three teams. These recordings covered hundreds of diverse exchanges, sufficient to demonstrate *ROGER* and gather feedback. We cleaned the audio records to remove background noise, and transcripts were created. We then computed fundamental and emotional voice metrics using the iMotions software [17]. Finally, all time-series data was synchronized to ensure alignment and eliminate any minor discrepancies.

We abstracted the data related to one training session in Table 1.

Data	Description
Participant	Each trainee is assigned a unique ID for tracking throughout the session. Optionally, participant names can be included for identification.
Voice records	Each participant's voice communication is captured as an individual, cleaned audio recording (MP4 format). Additionally, we generated a composite recording of all participants.
Heart rate variability (HRV)	We computed the root mean square of successive differences between normal heartbeats (RMSSD), a heart rate variability (HRV) metric, to assess acute mental stress. RMSSD has proven to be a reliable indicator of stress, even when measured over short time intervals [30]. This was done for a moving time window of 30 seconds for every second of the training, to enable a real-time metric for acute stress, adapted from Zechner et al. [50].
Voice metrics	Using the iMotions platform with audEERING, we analyzed voice records to extract arousal, dominance, and valence metrics. audEERING is an AI-based audio analysis tool. Through this integration, we assessed participants' emotional expressions based on these key dimensions.
Transcript	Tabular data with one message per row. Columns: ID, start (seconds), end (seconds), text, trainee ID.
Contextual information	Event data provided as a sequence of timestamps and text labels. The (VR) video feeds are also considered contextual data.

Table 1: Raw and derived data used in *ROGER*.

6 Conceptual Motif Model

We propose a high-level conceptual model for analyzing communication patterns, called **motifs**. Given the lack of standardized guidelines, the objective of the model is to accommodate varying training objectives and individual analysis needs.

The term *motif* is frequently used in graph analysis, where it refers to a fundamental pattern [48] or building block [24] of a complex network. We adopt this general idea for communication to flexibly analyze key patterns and characteristics in verbal exchanges. The model serves as a foundation for examining essential aspects of verbal communication and is implemented as a proof-of-concept motif-querying feature in our interface, as described in Sec. 7.5.

Based on the analysis of domain needs, we identified five key categories of motifs: **Content-Based Motifs** focus on the spoken content of communication and help to address questions such as *Are the messages brief, clear, and essential?* or *Are they using the learned terminology?* **Language Style Motifs** capture variations in the manner of speech, including the use of direct orders, excessive filler words, and sentence complexity. **Structural Motifs** help to analyze the flow and organization of communication, such as turn-taking dynamics, interruptions between colleagues, and the sequencing of interactions. **Contextual Motifs** relate to the situational context of communication, such as external events or locations. **Emotional and Stress-Related Motifs** help to analyze how emotions and stress influence the effectiveness and overall dynamics of communication.

Motifs can be applied individually or in combination, depending on analysis needs. An individual motif targets a specific aspect of communication (e.g., identifying critical keywords or analyzing event-related messages). In contrast, combined motifs allow a multifaceted analysis (e.g., combining content-based motifs with emotional motifs allows assessing how stress contributes to failures in following standardized protocols). This allows the analysis of either isolated communication elements or broader, interconnected patterns. Motif application can follow either inclusive or exclusive criteria. Inclusive criteria select any message that matches one or more specified motifs while exclusive criteria only select messages that simultaneously match all specified motifs.

Our model provides a versatile foundation for an AI-based analysis of communication patterns through four key querying actions [26]: filter, highlight, summarize, and compare. *Filter* and *highlight* help to identify individual messages, while *summarize* aggregates messages that match a motif, such as incomprehensible communication per speaker. *Comparison* supports side-by-side analysis of motifs. Although the feedback we collected did not indicate a strong demand for this functionality, it is included for completeness.

7 *ROGER*: Interactive Visualization Interface

We now describe the overall user interface, its visualizations, and interaction techniques, and discuss the design rationales with respect to the data abstractions and domain challenges.

7.1 Overview

The interface, shown in Fig. 1, features carefully designed linkages between components. The main visualization of voice metrics is in the center, with the option to switch between summary charts and a detailed communication timeline. This option addresses *C7: Limited Analysis Time* and enables users to quickly gain an overview while supporting more in-depth investigations. The transcript of communication is displayed alongside the visualizations, either as a *conversation log* in the right sidebar or as a *conversation cascade* at the bottom (see Sec. 7.4).

The header section includes audio and video controls (see Sec. 7.2) and various filtering options, such as participants, time range, and custom search queries which are applied across all views. Users can see if motifs have already been applied and can open a pop-up panel to create or select new motifs (see Sec. 7.5).

We use categorical colors to distinguish between team members, as previous work shows that up to seven color categories are well distinguishable when luminance is kept consistent [13], aligning with the typical maximum team size. All other data attributes are encoded using alternative visual channels.

7.2 Audio and Video Controls

This component allows users to replay audio for all participants simultaneously or for individuals (*C6: Recall Difficulties*). Voice recordings are commonly used in training reviews, particularly in law enforcement and emergency response. To facilitate the transition to our approach, we retain this familiar practice while extending it with visualizations, annotations, and AI-based filters. These additions are intended to support a more structured assessment of tone and message delivery (Task T2).

Optionally, a separate window for video replay can be opened to provide a corresponding video feed from the training. The aim is to analyze communication within its broader context, rather than in isolation (*C5: Contextual Analysis*), such as identifying failures like *"Officer turns away from a potential danger while speaking"*. Given the limited screen space already occupied by multiple coordinated views, we found it best to use a second screen which is typically available for after-action reviews.

7.3 Communication Timeline

The primary visualization, shown in Fig. 1 in the center, presents the communication over time (left to right). We describe the component from top to bottom.

At the top, below the time-range filter, we display the **scrubber**, which moves automatically during playback, and users can jump to specific moments by manually dragging along the x-axis; a common interaction in video and audio editing tools [34].

Event timestamps, shown as diamond shapes, provide contextual orientation (*C5: Contextual Analysis*). In the reported policing scenarios, events always occur as distinct points in time, but the interface could be extended to support time intervals too.

The **streamgraph** visualizes the speaking share over time; a commonly used technique to visualize multiple time series [5, 6]. We construct it by binning the raw transcript based on a predefined bin size (9 seconds by default; interactively adjustable via an input slider). When zoomed out, the streamgraph reveals the overall shape of the conversation (e.g., to address *C1: Diverse backgrounds* and *C2: Stress and emotions*). It helps answer questions related to Task T3, such as: *Does most of the conversation happen at the beginning or end? Are certain team members more present? Does the speaking activity shift after key events?*

Below the streamgraph, we display the **speech activity bands**, with each row corresponding to a team member. In the foreground, the speech bubbles indicate when and for how long a person was speaking. This representation is intended to support more detailed per-person observations and complement the streamgraph. In the background, we represent stress levels using a gradient from white (no stress) to dark grey (high stress). The design rationale was to emphasize general stress patterns over time (*C2: Impact of stress and emotions*) while acknowledging the inherent uncertainty in fine-grained stress measurements [30].

Initially, we experimented with embedding transcript messages directly into the speech bubbles and developed an algorithm to adjust their vertical positions within the bands to avoid overlaps. We also tested transposing the entire timeline to depict time from top to bottom to align with the transcript's vertical scrolling. However, through stakeholder feedback, we learned that it is most intuitive for users if the visualization provides an overview of the conversation, with time depicted from left to right, while the transcript is presented in a separate panel with top-to-bottom scrolling, as described in the following section.

7.4 Conversation Log & Cascade

Users can choose to view the transcript either as a chronological list of messages in the conversation log (see Fig. 1, right) or as a conversation cascade (see Fig. 2), where messages are grouped by team member. These components support Task T1 by enabling the analysis of spoken content. In both views, messages are displayed with the member's name, message content, and, if enabled, emotional characteristics using our proposed SAM Glyphs (see Sec. 7.7 for further details). The ability to analyze messages in detail directly addresses *C4: Fuzzy guidelines*. Clicking on a message moves the scrubber and plays the corresponding audio recording.

The conversation log presents a familiar, compact format for reviewing chat logs, where each row represents a message. While space-efficient, this format hides certain dynamics, such as parallel conversations or pauses. In contrast, the conversation cascade arranges messages in separate columns for each team member, inspired by Flow Client [45] and Chat Spaces [11], which address non-linearity and the loss of time-sequencing information. Messages are aligned vertically by timestamp, but to prevent overlap, they are shifted downward when they collide. Large gaps where no exchanges occur are removed to keep the visualization compact.

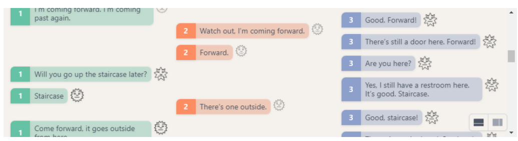

Figure 2: Excerpt of the conversation cascade with three team members.

7.5 Motif-Querying Feature

We implemented a motif-querying feature based on the proposed conceptual model (Sec. 6) to support a more flexible, intelligent analysis of communication patterns. This feature is primarily motivated by *C4: Fuzzy Guidelines*. Some users focus on inspecting specific phrases and keywords, while others need to analyze broader conversation dynamics. When users open the pop-up panel, they can choose from predefined motifs, such as identifying all messages that lead to high stress or messages with excessive use of filler words. Additionally, they can create custom LLM-based motifs through a user-defined prompt. In essence, motifs allow users to select messages from the transcript for identification (highlighting or filtering) and summarization (e.g., count all message instances per team member that match the motif) tasks.

As a proof-of-concept, we implemented a set of predefined motifs based on questions articulated during the design process. Structural, contextual, and stress-related motifs are computed using the available datasets. For content-based and language-style motifs, we leverage an LLM (Claude 3.5 Sonnet). The prompt template is included in the supplemental material.

7.6 Contextual Conversation Panel

A context panel with an embedded conversation log can be opened by clicking on bands in the streamgraph or segments in the stacked bar chart. We learned that stacked bar charts are particularly effective for providing an at-a-glance overview (*C7: Limited analysis time*), such as displaying stress levels for each team member. However, follow-up questions often arose regarding specific content (e.g., high-stress messages). The context panel addresses this by allowing users to scroll through messages that match selected criteria.

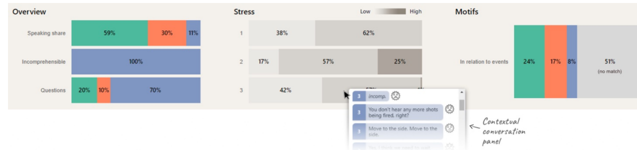

Figure 3: Stacked bar charts with relative message counts for each team member. The user views messages from team member #3 with stress level 2 (on a scale from 1-4) in the contextual conversation panel.

7.7 Affective State Glyphs (SAM Glyphs)

To visualize the emotional characteristics of messages (*C2: Stress, emotions*), we developed SAM glyphs, a compact representation based on the Self-Assessment Manikin (SAM) model [3]. SAM is a widely-used pictorial assessment technique to measure the three primary emotional dimensions: valence, arousal, and dominance. Due to limited screen space, we condensed these three data attributes into a single visual element to create an integrated, space-efficient encoding. An overview of various states is presented in Fig. 4. Arousal is depicted using a circle that transitions from smooth (calm) to spiky (excited). Valence is encoded through facial expressions, ranging from a sad face to a happy face. Dominance is presented through both the size and opacity of the glyph: larger, more opaque glyphs indicate higher dominance. Notably, color encoding was avoided because it is already used to represent team members in the interface.

Figure 4: SAM glyphs are a compact visual representation of emotional characteristics inspired by the Self-Assessment Manikin [3].

8 Preliminary Evaluation and Design Implications

We conducted a qualitative analysis of the interviews and workshop discussions. The analysis followed an inductive coding approach [4], where themes were derived directly from the participants' statements. This process involved multiple rounds of refinement to ensure the themes accurately capture the range of topics mentioned by participants, without imposing external assumptions. The following themes reflect the participants' perspectives on *ROGER* and training needs.

Capturing and visualizing verbal communication. Participants emphasized the potential of the interface to address a wide range of analysis needs. Communication failures were noted to have a significant impact on overall performance. It was repeatedly highlighted that *ROGER* allows them to observe general communication dynamics while also providing access to the transcript with individual messages. Medical experts noted the cross-domain applicability in our adaptation workshop: *"Having an after-action review tool for* *communication would be novel"*; *"I think this would be excellent for resuscitation training."*

Diverse analysis needs. Analysis requirements vary significantly depending on the user's role and the type of training. Some interviewees emphasized the importance of high-level summaries of communication, while others preferred detailed analysis, such as examining stress impact on communication or verifying the use of new terminology. This range of needs highlights the demand for adaptable and intelligent analysis tools.

Integration into training routines. Participants stressed that new technologies or analysis tools must be introduced without disrupting existing training routines or scenarios. Major changes to the training structure could raise concerns about additional workload or complexity. Across all interviews and workshops, participants emphasized the importance of an automated pipeline that enables *ROGER* to be used immediately after a training session.

Motif-based querying. Participants found motif-based querying very compelling. Interviewee P5 mentioned, *"This features makes it flexible enough to be used for detailed communication training, not just for tactical after-action reviews."* LLM queries generally worked well, although in a few cases, irrelevant messages were included in the results. However, the tool was praised for helping to filter a large set of messages into a manageable subset for further analysis.

Nonverbal Cues and Contextual Factors. Communication should not be analyzed in isolation, as frequently noted. Certain nonverbal cues remain difficult to capture, even with a secondary video feed. A common example in the police context is team members turning toward colleagues and away from potential danger when speaking.

Physiological measurements. Correlating heart rate variability measurements with communication patterns was seen as highly relevant, as stress simulation is a core component of many training exercises. However, some participants expressed skepticism about the reliability of these measurements. Playing the corresponding voice recordings and reading the transcripts helped assess the relationship between stress levels and communication behavior.

9 Discussion

Our findings suggest that *ROGER* can support a comprehensive analysis of team communication, particularly through the interplay of motif querying and interactive visualization. We reflect on design and methodological decisions in *ROGER*, and discuss its limitations.

9.1 Reflections on the Communication Scope

Our approach focuses on inward communication within teams of up to 7 members. While this has proven valuable for analyzing intra-team dynamics, it presents a limitation in addressing outward communication with other teams or external parties. In *ROGER*, we can consider the control room as part of the team, but for certain analyses, it may be beneficial to visually distinguish between team members on the ground and other colleagues.

Training communication with external parties is crucial in law enforcement, particularly for de-escalation [49]. But it is also relevant in other domains where teams frequently interact with the public. While there may be some overlap with inward communication patterns, addressing outward exchanges requires distinct functionalities and design considerations [19]. This makes it an important direction for future work.

9.2 Reflections on the LLM-Based Motifs

The use of LLMs to sift through transcripts and identifying messages has proven to be highly effective, compared to more traditional search methods. We leverage LLMs for content-based and language-style motifs solely based on the transcript. Our findings suggest that LLMs support this use case, but their effectiveness depends on motif complexity. More intricate motifs require precisely formulated instructions to achieve reliable results (i.e., prompt engineering). A potential future direction is to explore user interface mechanisms that guide users in optimizing motif definitions.

9.3 Reflections on the Methodology

We followed the design study methodology [35], starting with understanding the problem context and narrowing our focus to a small group of stakeholders. Police trainers supported our intuition that other fields might have similar analysis needs, but we only reached out to another domain later in the process. Medical experts confirmed our assumptions and the strong overlap in challenges and analysis questions with policing. In hindsight, it would have been beneficial to engage multiple domains earlier to discuss mockups and gather more diverse feedback.

Evaluating visualization systems for real-world applications presents unique challenges, especially when no comparable baseline tools exist and the analysis itself is inherently fuzzy and unstructured. Existing training methods involve manual review of voice recordings to identify key moments, but no established baseline provides comparable analysis techniques. Thus, we collected feedback across multiple iterations and from different organizations which aligns with established qualitative approaches where the goal is to develop a grounded understanding of a system's effectiveness [18]. We also conducted a summative adaptation workshop to assess cross-domain applicability. However, a more systematic evaluation is necessary to assess its long-term impact in training environments, capture both qualitative and quantitative insights, and examine its adoption by organizations.

10 Conclusion

In this paper, we introduced *ROGER*, an interactive visualization system designed to support after-action reviews of team communication in high-stress environments. By integrating a flexible, LLM-powered motif-querying feature, *ROGER* facilitates a comprehensive exploration of communication patterns. It supports the analysis of content, delivery, and team dynamics, while accounting for contextual factors, emotions, and stress levels.

Our preliminary findings from workshops and interviews with police trainers and medical first responders indicate potentially significant societal benefits. For example, improved communication may reduce conflict escalation and enhance public safety. In emergency medical contexts, accurate information flow is essential for patient care and decision-making. Verbal communication is also critical in other high-pressure environments, such as fire brigades, rescue services, and control rooms, where efficient coordination can directly impact outcomes.

Acknowledgments

This research was supported by the III.Lab (Intelligent Interfaces Innovation), funded by AIT Austrian Institute of Technology GmbH, University of Salzburg and Land Salzburg (GA No.20102/F2300416-KZP). ChatGPT assisted in refining and condensing some paragraphs, with validity checked by the authors.

References

[1] Chris Baber, Andrew Leggett, Simon Attfield, and Emmeline Elliott. 2022. Using Speech Act Theory to Apply Automated Communications Analysis to Distributed Sensemaking. In *Proc. of the Human Factors and Ergonomics Society*, Vol. 66. 55–59.

[2] Tony Bergstrom and Karrie Karahalios. 2007. Conversation Clock: Visualizing audio patterns in co-located groups. In *Int. Conf. System Sciences (HICSS*. 78–78.

[3] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59.

[4] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3 (2021), 328–352.

[5] Chuan Bu, Quanjie Zhang, Qianwen Wang, Jian Zhang, Michael Sedlmair, and et al. 2020. Sinestream: Improving the readability of streamgraphs by minimizing sine illusion effects. *IEEE TVCG* 27, 2 (2020), 1634–1643.

[6] Lee Byron and Martin Wattenberg. 2008. Stacked graphs–geometry & aesthetics. *IEEE TVCG* 14, 6 (2008), 1245–1252.

[7] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-time capture and visualization of verbal content in meetings. In *Proc. ACM Conf. on Human Factors in Computing Systems (CHI)*. 1–14.

[8] Hannah Davis, Sophie Lawrence, Emily Wilson, Fay Sweeting, and Annabel Poate-Joyner. 2023. 'No one likes a grass' Female police officers' experience of workplace sexual harassment: A qualitative study. *Int. Journal of Police Science & Management* 25, 2 (2023), 183–195.

[9] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel Keim. 2016. ConToVi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum* 35, 3 (2016), 431–440.

[10] Ehsan Garosi, Reza Kalantari, Ahmad Zanjirani Farahani, Mojgan Zuaktafi, Esmaeil Hosseinzadeh Roknabadi, and Ehsan Bakhshi. 2020. Concerns about verbal communication in the operating room. *Human Factors* 62, 6 (2020), 940–953.

[11] Werner Geyer, Andrew J Witt, Eric Wilcox, Michael Muller, Bernard Kerr, Beth Brownholtz, and David R Millen. 2004. Chat spaces. In *Proc. ACM Conf. on Designing Interactive Systems (DIS)*. 333–336.

[12] Susan Havre, Elizabeth Hetzler, Paul Whitney, and Lucy Nowell. 2002. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE TVCG* 8, 1 (2002), 9–20.

[13] Christopher G Healey. 1996. Choosing effective colours for data visualization. In *Proc. IEEE Visualization*. 263–270.

[14] Kelly A. Hine and Adelaide Bragias. 2021. Effective communication during major crises: a systematic literature review to identify best practices for police. *Police Practice and Research* 22, 5 (2021), 1492–1507.

[15] Kelly A Hine and Adelaide Bragias. 2021. Effective communication during major crises: a systematic literature review to identify best practices for police. *Police Practice and Research* 22, 5 (2021), 1492–1507.

[16] Enamul Hoque and Giuseppe Carenini. 2014. ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum* 33, 3 (2014), 221–230.

[17] iMotions. 2025. iMotions Voice Analysis. https://imotions.com Acc.: Jun 2025.

[18] Tobias Isenberg, Petra Isenberg, Jian Chen, Michael Sedlmair, and Torsten Möller. 2013. A systematic review on the practice of evaluating visualization. *IEEE TVCG* 19, 12 (2013), 2818–2827.

[19] Yasmeen Krameddine and Peter H Silverstone. 2015. How to improve interactions between police and the mentally ill. *Frontiers in Psychiatry* 5 (2015), 186.

[20] Xingyu Lan, Yanqiu Wu, and Nan Cao. 2023. Affective visualization design: Leveraging the emotional impact of data. *IEEE TVCG* (2023).

[21] Henrietta Lee, Robyn Woodward-Kron, Alan Merry, and Jennifer Weller. 2023. Emotions and team communication in the operating room: a scoping review. *Medical Education Online* 28, 1 (2023), 2194508.

[22] Jialang Victor Li, Max Kreminski, Sean M Fernandes, Anya Osborne, Joshua McVeigh-Schultz, and Katherine Isbister. 2022. Conversation balance: A shared VR visualization to support turn-taking in meetings. In *CHI Conf. on Human Factors in Computing Systems (Ext. Abstracts)*. ACM, 1–4.

[23] Pooja Mathur and Karrie Karahalios. 2009. Visualizing remote voice conversations. In *CHI Conf. on Human Factors in Computing Systems (Ext. Abstracts)*. 4675–4680.

[24] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. 2002. Network motifs: simple building blocks of complex networks. *Science* 298, 5594 (2002), 824–827.

[25] Wookhee Min, Randall Spain, Jason D Saville, Bradford Mott, Keith Brawner, Joan Johnston, and James Lester. 2021. Multidimensional team communication modeling for adaptive team training. In *Int. Conf. on Artificial Intelligence in Education*. 293–305.

[26] Tamara Munzner. 2014. *Visualization analysis and design*. CRC Press.

[27] Markus Murtinger, Jakob Carl Uhl, Lisa Maria Atzmüller, Georg Regal, and Michael Roither. 2024. Sound of the Police—Virtual Reality Training for Police Communication for High-Stress Operations. *Multimodal Technologies and Interaction* 8, 6 (2024), 46.

[28] Jay Pande, Wookhee Min, Randall D Spain, Jason D Saville, and James Lester. 2023. Robust team communication analytics with transformer-based dialogue modeling. In *Int. Conf. on Artificial Intelligence in Education*. 639–650.

[29] Refense. 2025. VR Training System. https://refense.com Accessed: Jun 2025.

[30] Lizawati Salahuddin, Jaegeol Cho, Myeong Gi Jeong, and Desok Kim. 2007. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *Int. Conf. Engineering in Medicine and Biology Society*. 4656–4659.

[31] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölinder. 2010. Mind the body! Designing a mobile stress management application encouraging personal reflection. In *Proc. ACM Conf. on Designing Interactive Systems (DIS)*. 47–56.

[32] Jason Saville, Randall Spain, Joan Johnston, and James Lester. 2022. An analysis of squad communication behaviors during a field-training exercise to support tactical decision making. *Proc. AHFE Int. Conf. Human Factors and Simulation* 30 (2022), 109–116.

[33] J. Saville, R. Spain, J. Johnston, and J. Lester. 2022. An Analysis of Squad Communication Behaviors during a Field-Training Exercise to Support Tactical Decision Making. In *Human Factors and Simulation*, Julia Wright and Daniel Barber (Eds.). Vol. 30. AHFE Open Access.

[34] Klaus Schoeffmann, Marco A Hudelist, and Jochen Huber. 2015. Video interaction tools: A survey of recent work. *Comput. Surveys* 48, 1 (2015), 1–34.

[35] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG* 18, 12 (2012), 2431–2440.

[36] Yang Shi, Chris Bryan, Sridatt Bhamidipati, Ying Zhao, Yaoxue Zhang, and Kwan-Liu Ma. 2018. MeetingVis: Visual narratives to assist in recalling meeting context. *IEEE TVCG* 24, 6 (2018), 1918–1929.

[37] Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. 2017. IdeaWall: Improving creative collaboration through combinatorial visual stimuli. In *Proc. ACM CSCW*. 594–603.

[38] Jaime Snyder, Mark Matthews, Jacqueline Chien, Pamara F Chang, Emily Sun, Saeed Abdullah, and Geri Gay. 2015. Moodlight: Exploring personal and social implications of ambient display of biosensor data. In *Proc. ACM Conf. on Computer Supported Cooperative Work & Social Computing (CSCW)*. 143–153.

[39] Anna Ståhl, Jonas Löwgren, and Kristina Höök. 2014. Evocative balance: Designing for interactional empowerment. *Int. Journal of Design* 8, 1 (2014), 43–57.

[40] Danique Stappers and Jun Hu. 2024. VIT. IN: Visualizing Collective Stress with a Dynamic Painting. In *Int. Conf. on Human-Computer Interaction*. 96–115.

[41] Simone Stroomer and Herre van Oostendorp. 2002. Analyzing Communication in Team Tasks. In *Cognition in a Digital World*. Psychology Press.

[42] Judith Tiferes and Ann M Bisantz. 2018. The impact of team characteristics and context on team communication: An integrative literature review. *Applied Ergonomics* 68 (2018), 146–159.

[43] Jakob C Uhl, Markus Murtinger, Olivia Zechner, and Manfred Tscheligi. 2022. Threat assessment in police VR training: multi-sensory cues for situation awareness. In *IEEE Int. Conf. on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. 432–437.

[44] Floris van den Oever and Jan Maarten Schraagen. 2021. Team communication patterns in critical situations. *Journal of Cognitive Engineering and Decision Making* 15, 1 (2021), 28–51.

[45] David Vronay, Marc Smith, and Steven Drucker. 1999. Alternative interfaces for chat. In *Proc. Symp. User Interface Software and Technology*. 19–26.

[46] Jiaqi Wang, Tianyi Gui, Mingzhi Cheng, Xuan Wu, Ruolin Ruan, and Meng Du. 2023. A survey on emotional visualization and visual analysis. *Journal of Visualization* 26, 1 (2023), 177–198.

[47] Mengru Xue, J Liang, Jun Hu, and LMG Feijs. 2017. ClockViz: designing public visualization for coping with collective stress in teamwork. In *Int. Conf. on Design and Semantics of Form and Movement (DeSForM)*. 67–78.

[48] Ömer Nebil Yaveroğlu, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, and et al. 2014. Revealing the hidden language of complex networks. *Scientific Reports* 4, 1 (2014), 4547.

[49] Benjamin Zaiser and Mario S Staller. 2015. The Word is Sometimes Mightier Than the Sword: Rethinking Communication Skills to Enhance Officer Safety. *Journal of Law Enforcement* 4, 5 (2015).

[50] Olivia Zechner, Helmut Schrom-Feiertag, Jakob Uhl, Quynh Nguyen, Lisanne Kleygrewe, and Manfred Tscheligi. 2023. Mind the Heart: Designing a Stress Dashboard Based on Physiological Data for Training Highly Stressful Situations in Virtual Reality. In *IFIP Conf. on Human-Computer Interaction*. 209–230.