Beyond Words: Decoding Facial Expression Dynamics in Motivational Interviewing

Nezih Younsi, Catherine Pelachaud, Laurence Chaby

Beyond Words: Decoding Facial Expression Dynamics in Motivational Interviewing

Nezih Younsi, Catherine Pelachaud, Laurence Chaby
ISIR, Sorbonne University
75005, Paris, France
younsi@isir.upmc.fr, catherine.pelachaud@isir.upmc.fr, laurence.chaby@isir.upmc.fr

Abstract

This paper focuses on studying the facial expressions of both client and therapist in the context of Motivational Interviewing (MI). The annotation system Motivational Interview Skill Code MISC defines three types of talk, namely sustain, change, and neutral for the client and information, question or reflection for the therapist. Most studies on MI look at the verbal modality. Our research aims to understand the variation and dynamics of facial expressions of both interlocutors over a counseling session. We apply a sequence mining algorithm to identify categories of facial expressions for each type. Using co-occurrence analysis, we derive the correlation between the facial expressions and the different types of talk, as well as the interplay between interlocutors' expressions.

Keywords: Non-verbal behaviour, Sequence Mining, Motivational interviewing

1. Introduction

Motivational Interviewing (MI) is a directive, client-centered therapeutic approach designed to collaboratively facilitate behaviour change by enhancing individuals' intrinsic motivation (Miller and Rollnick, 2012), which represents an individual's desire to initiate change, driven by personal values and beliefs (Vallerand, 2000). MI's efficacy is the the ability to foster meaningful client-therapist interactions, serving both as a mechanism for establishing rapport (Deci and Ryan, 2012; Van Minkelen et al., 2020) and a medium for targeted interventions. Both verbal and non-verbal cues are pivotal in this context, serving as tools to enhance the quality of interaction and rapport-building (Tickle-Degnen and Rosenthal, 1990).

Within the MI framework, several authors have created dataset based on real motivational interviews, containing series of counseling videos, annotated following a specific MI code scheme. Coding frameworks like MISC (Motivational Interviewing Skill Code) (Miller et al., 2003) and MITI (Motivational Interviewing Treatment Integrity) (Moyers et al., 2003) are frequently employed to identify MI codes and behaviours associated with both therapist and client. These annotations establish a structured guideline, categorizing therapist behaviours such as Reflection, Question or Advice and, client type of talks such as Change, Sustain, and Neutral Talk. These categorizations facilitate the evaluation of interactions within MI from researcher's desired perspective. While the verbal aspects of MI have been extensively studied, there is a notable gap in the literature concerning the role of non-verbal behavior, specifically facial expressions, in such interactive contexts (Torre et al., 2021), where they play a key role in establishing a social rapport (Tickle-Degnen and Rosenthal, 1990) and facilitating intrinsic motivation initiation during MI sessions.

Our aim is to explore the co-occurrences between distinct categories of MI type of talks and the facial expressions exhibited by both therapists and clients. The study also extends to analyze the reciprocal interplay between both interlocutors facial expressions during MI sessions. By focusing on these under-investigated components of MI, the current research aims to augment the extant literature, thereby offering a more comprehensive understanding of non-verbal behavior in Motivational Interviewing.

2. Background

Motivational Interviewing (MI) is a client-centered counseling approach that seeks to amplify an individual's intrinsic motivation to behaviour change by addressing and resolving ambivalence. At the heart of MI is the concept of motivation, which spans a continuum from extrinsic to intrinsic motivation. While extrinsic motivation comes from external rewards and praise, intrinsic motivation comes from an internal desire to accomplish a goal (Vallerand, 2000). The primary objective of MI is to initiate intrinsic motivation, with a particular emphasis on fostering autonomy. Autonomy, defined as the capacity to make choices and dictate one's actions, is a pivotal element that therapists leverage during MI sessions to initiate intrinsic motivation (Resnicow et al., 2002; Miller and Rollnick, 2012). This emphasis on autonomy has led researchers to draw parallels between MI and Self-Determination Theory (SDT) (Deci and Ryan,

LREC-COLING 2024, pages 2365–2374

2012). SDT posits three fundamental psychological needs: autonomy, competence (the belief in one's ability to learn and gain skills), and relatedness (the sense of connection and belonging to a group). Several studies have attempted to explore the interplay between MI and the components of SDT. For example, (Van Minkelen et al., 2020) explored how relatedness, plays a role in motivational therapy settings, finding a strong link between non-verbal signals and this sense of connection. (Baker et al., 2020) supported this, highlighting a tie between relatedness and social rapport, which is was also defined as the sense of connection and harmony felt during a a social interaction by (Tickle-Degnen and Rosenthal, 1990). This social rapport is built through positive emotions, shared attention, and coordinated actions over an interaction. Additionally (Torre et al., 2021) recently pointed out the importance of smiling as a non-verbal signal in therapy sessions, underscoring its part in expressing positive emotions, a key element in building social rapport.

Building on this foundation, he interplay between verbal and non-verbal cues materializes as a critical determinant of the quality of the social interaction. The nuances of these behaviours can significantly influence the dynamics of therapeutic interactions from a social standpoint (Burgoon et al., 1995; Klohnen and Luo, 2003; Berscheid, 1994). These dynamics can be conceptualized as adaptations of interpersonal verbal and non-verbal behaviours. Given the theoretical considerations of Self-Determination Theory and the theory of rapport as guiding frameworks, it becomes relevant that non-verbal behavior adaptation holds significant potential in enhancing the interaction quality of Motivational Interviewing.

3. Related Works

In the computer science domain, MI has elicited considerable interest from researchers. A particular effort is being made in exploiting annotated corpus, obtain from public and private data sources, presenting interactions between humans in MI contexts (Wu et al., 2023; Pérez-Rosas et al., 2016; Rubak et al., 2005). The use of computational methodologies aims to gain a better understanding of behavioural specificities and their consequent implication in therapeutic session.

Several researches in MI have been dedicated to the development of automated tools for counseling practices quality assessment. (Pérez-Rosas et al., 2019) investigated the linguistic dynamics inherent in counseling conversations. Her analysis of interaction patterns, linguistic alignment, sentiment, and thematic content unveiled distinctive linguistic markers that differentiate high-quality counseling from its lower-quality counterpart. In parallel, (Lord et al., 2015) analyzed language style synchrony between counselors and clients, relying on the Linguistic Inquiry and Word Count (LIWC) lexicon to measure how counselors match their clients' language. Additionally, in terms of automatic behaviour assessment and detection, (Xiao et al., 2012) integrated acoustic and linguistic data to assess the empathetic responses of counselors.

In the recent years, with the advances of machine learning (ML), studies started using ML models and Natural language processing (NLP) techniques to optimize these detection and evaluation techniques on MI corpus. Recent contribution by (Tran et al., 2023) out-passes the previous works concerning the assessment of therapist empathy during MI sessions, using machine learning techniques and verbal features to provide insights into the therapeutic process. (Tanana et al., 2015) utilized recursive neural networks to identify counselor statements that discuss client change talk. However, a common thread weaving through these studies is their predominant focus on verbal and textual data. Except the notable work done by (Nakano et al., 2022), which accentuates the role of both verbal and facial cues in discerning change talk during MI, focusing on the client's perspective and specific behaviours.

A gap in the existing literature is the limited attention to non-verbal cues in the context of MI. Based on our theoretical framework and using sequence mining algorithms, our study is positioned to bridge this gap, aiming to unravel the intricate dynamics of non-verbal behaviours and their influence on the quality of MI sessions. We are particularly interested in exploring the co-occurrences between specific MI behaviours (Type of talks) and the non-verbal cues, specifically the facial expressions exhibited by both therapists and clients. The study also extends to analyze the reciprocal interplay between both interlocutors facial expressions during MI sessions. Given this backdrop, our work aims to answer to these specific research questions.

- RQ1 : Do specific MI behaviours or annotated types of talk co-occur with distinct categories of facial expressions?
- RQ2 : Do therapists exhibit adaptive behaviours in response to clients that positively influence MI quality?
- RQ3 : How significant is positivity management, particularly through positive facial expressions, in the quality of MI?

4. Methodology

The methodology section starts with an overview of the AnnoMI database which serves as our primary data source. Next, it describes the data extraction, filtering, and preprocessing steps, necessary for preparing the dataset for the computational analysis. The section concludes with a description of the sequence mining techniques used to study co-occurrences between dialogue types and facial action units.

4.1. AnnoMI data base

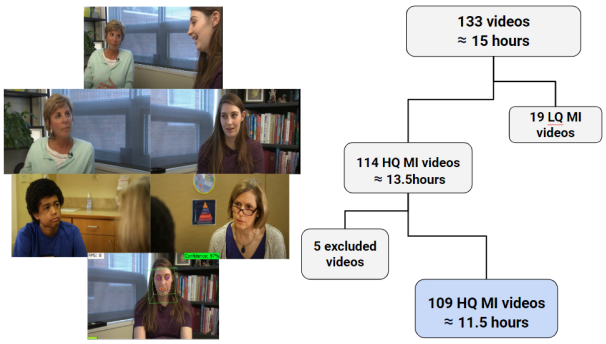

Figure 1: 133 faithfully transcribed and expert-annotated demonstrations of high quality (HQ) and low-quality (LQ) motivational interviewing (MI)

The AnnoMI database, comprises 133 transcribed therapy conversations collected from YouTube and Vimeo, each illustrating key MI techniques (Wu et al., 2023). These sessions were closely annotated by experts in MI using the Motivational Interviewing Skill Code (MISC) (Miller et al., 2003). For therapists, behaviours are grouped into: Question, Input, and Reflection. Questions are either open-ended or closed. Inputs cover knowledge sharing, advice, options, and goal-setting. Reflections, vital in MI, are either simple or complex, with the latter adding deeper meaning to client statements.

Client annotations are based on their stance towards change: Change Talk (favoring change), Sustain Talk (resisting change), and Neutral Talk (no clear preference). This categorization, rooted in MI coding standards and therapist insights, facilitates a comprehensive analysis. The AnnoMI dataset, with its verbal annotations and video data, is crucial for our exploration of verbal and non-verbal cues in MI, emphasizing the significance of facial expressions in MI effectiveness.

To ensure the quality of our study, we chose to include only high-quality MI videos. According to the literature (Miller and Rollnick, 2012), in high-quality MI, therapist behaviour is more empathetic and client centered, whereas low-quality MI is characterised by objective instructions and suggestions. This decision aligns with our focus on examining the appropriate non-verbal behaviours that enhance the quality interaction of MI, rather than those that may detract from it. Additionally, we excluded five videos from the dataset due to their low video quality and unfavorable filming angles, ensuring that our analysis is based on the most reliable and relevant. In summary, we started with 133 video interactions of both high and low-quality MI, totaling 15 hours of counseling sessions. After this filtering, we were left with 109 high-quality counseling videos, amounting to 11.5 hours, featuring 109 distinct therapist-client pairs (Figure 1).

4.2. Extraction of Action Units

In the initial phase, we employed OpenFace (Baltrusaitis et al., 2018), a computer vision toolkit proficient in facial landmark detection, head-pose estimation, and eye-gaze estimation. OpenFace also outputs facial action units (AUs) (Du et al., 2014), which are based on the Facial Action Coding System (FACS) (Ekman and Friesen, 1978) to encode actions of grouped muscles involved in facial expressions. Each video interaction in the AnnoMI dataset was represented by three files: one for the client's OpenFace data, another for the therapist's OpenFace data, both captured at a rate of 25 frames per second (25 Hz), and a third file containing the AnnoMI transcripts and their MISC annotations at the utterance level.
the OpenFace files of both client and the therapist were combined into a single file. From this merged dataset, we specifically extracted the Action Units (AUs) while omitting other features. This resulted in 2 x 18 facial features, with 18 for each client and therapist. Each action unit is represented by an activation feature and an intensity activation value. The data was then subjected to a median filter for noise reduction and underwent interpolation to fill in any missing values. Additionally, we set an intensity threshold of 0.5 for the activation threshold of action units to ensure the reliability of the data.

4.3. Synchronisation and chunking of Transcripts with OpenFace Data

Following the filtering of the extracted Action Units (AUs) from all videos, the next critical step was to synchronize these data with the transcript and annotation files. Given that the transcript was at the utterance level and the AUs were captured at a rate of 25 frames per second, synchronization based on video timestamps was essential to ensure accurate alignment and matching of dialogues and their MISC annotations with corresponding facial expressions.

To further refine the data, we employed a sequencing approach that chunked the data based

on silences and speaking turns. This method allowed us to construct sequences that corresponded to either a single speaking turn or a turn-taking silence. Each sequence was then characterized by several indicators: whether it was a silence or speaking phase, whether the client or therapist was speaking, and a series of events represented by the activated AUs during the chunk, along with the associated MISC code annotation for that turn as shown in Figure 2.
An "event" denotes a specific AU activation or a MISC-coded type of talk. For instance, an event might be the therapist's AU06 activation, linked to smiling, or a "Question" from the MISC code (Miller et al., 2003). Each event retained its start time and duration, preserving the sequence's temporal context and enhancing our understanding of interaction dynamics.

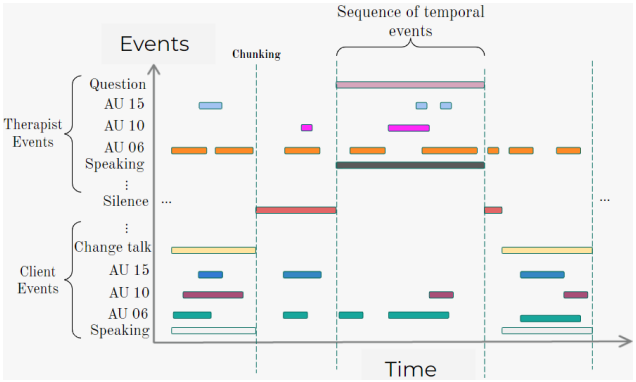

Figure 2: Representation of Temporal sequence chunking of data with event example

After chunking the data into sequences of events, the next step involved encoding these sequences into numerical formats. Each unique event was assigned a unique integer identifier descriptor. We ended up encoding the 109 videos and their annotations into a total of 22958 sequences, described by 49 possible events (the speaker, the AUs of both client and therapist, their MI behaviours and type of talks).

4.4. Dimension Reduction

Our purpose is to study co-occurrences between facial expressions events and MI behaviours. To do so, we decided to use a sequence mining algorithm that allows identifying recurrent event patterns in series of sequences. However, before diving into this analysis, it's crucial to address the high dimensionality of our data. Dimensionality reduction is essential in this context. An abundance of features may severely reduce algorithmic performance and increase processing time. By streamlining our dataset, we enhance efficiency and accuracy, focusing on the most pertinent features and eliminating potential redundancies.

At this stage of the study, our dataset was characterized by a multitude of features to describe a sequence: 2 x 18 facial features (both activation and intensity), speakers and silence descriptors , and 3 types of MISC talks for each of interlocutor (client and therapist). Given this extensive feature set, there was a need to pare down the data's complexity and dimensionality.

Initially, we chose not to consider the intensity of activation for the Action Units (AUs). We decided to only take into account the activation value (0 or 1) simplifying our dataset and enhancing the efficiency of our intended sequence mining algorithm. Also, isolated events with a duration shorter than 0.4 seconds were also disregarded, being viewed as potential noise.

To further refine our data, we employed hierarchical clustering using the City Block (Manhattan) distance metric (Nielsen and Nielsen, 2016). This method allowed us to amalgamate closely overlapping similar events allowing us to considerably reduce data complexity.

To manage the extensive number of events, we employed a support analysis approach. By calculating the occurrence proportion of each event within sequences, we discerned the most prevalent events:

$$P(e) = \frac{\text{Sequences containing the event } e}{\text{Total number of sequences}} \quad (1)$$

By setting a minimum occurrence threshold at 0.05, our intention was to filter out rare events, directing our attention to patterns that are more prevalent and potentially impactful. Yet, certain events surpassing this threshold were intentionally left out. Notably, $AU45$, which is frequently observed, was excluded due to its association with eye blinking. The consistent presence of this Action Unit could result in multitude of of patterns that, while recurrent, may not align with the core objectives of our research.

To reveal the co-occurrence patterns between the different events that constitute the sequences, our initial sequence mining approach was inspired by the HCApriori algorithm (Dermouche and Pelachaud, 2016), which combines hierarchical clustering with the Apriori sequence mining algorithm (Borgelt and Kruse, 2002). However, this method didn't yield satisfactory results. One of the primary challenges we encountered was the high variability in the duration of events, which made it difficult to identify consistent and meaningful occurrence patterns. Recognizing these challenges, we decided to adapt our approach. We preserved the order of appearance of events

within each sequence but chose to exclude the duration and starting time parameters. This decision was pivotal in simplifying our data structure, making it more amenable to sequence mining. Using the seq2pat algorithm (Wang Xin, 2022), which is based on the Constraint-based Sequential Pattern Mining (CSPM) approach (Chen and Hu, 2006), we adeptly identified patterns in sequences.

Upon analyzing the results, we observed high co-occurrence between specific AUs. For instance, $AU25$, $AU6$, and $AU12$ frequently co-occurred. This led us to group them functionally into broader categories:

- Mouth_Up Category: Comprising of $AU6$, $AU25$, and $AU12$.
- Nose_Wrinkle Category: Encompassing $AU9$ and $AU10$.
- Mouth_Down Category: Including $AU14$ and $AU15$.

We also observed that $AU26$ predominantly co-occurs during speaking intervals, being consistently present with speech. Given the high prevalence of this Action Unit, it tends to produce a multitude of patterns. While these patterns are frequent, they may not offer significant insights aligned with our research goals. To ensure a more targeted and meaningful analysis, we decided to exclude $AU26$. This exclusion allows us to hone in on patterns that are truly relevant and central to the objectives of our study.

With these new categories in place, we proceeded to apply hierarchical clustering once again. This time, our aim was to temporally fuse events within these categories, treating them as singular events. The rationale behind this was to reduce the granularity of our data, focusing on broader facial expression categories rather than individual AUs. Following this refinement, we re-applied the sequence mining analysis. This iterative approach, combining both functional understanding of facial expressions and data-driven insights, aimed to provide a more holistic understanding of the non-verbal cues in Motivational Interviewing.

5. Results and Discussion

Upon analyzing the refined data using the pattern analysis, numerous significant patterns of length 2 emerged. A closer examination of these patterns revealed consistent co-occurrences between specific types of talks and facial expression categories.

5.1. Type of talks and Categories of Facial expression

To assess the interplay between types of talks and facial expression categories, we compute a co-occurrence support matrix. This matrix was designed to encapsulate the support values representing the co-occurrence frequency between distinct events. Specifically, each matrix cell, denoted by $M_{i,j}$ represents the support value of co-occurrence between the event from row i and the event from column j. The matrix can be described as:

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{bmatrix}$$

Where:

- M_{ij} is the support value of co-occurrence between event i and event j.
- n is the total number of distinct events considered in the matrix.

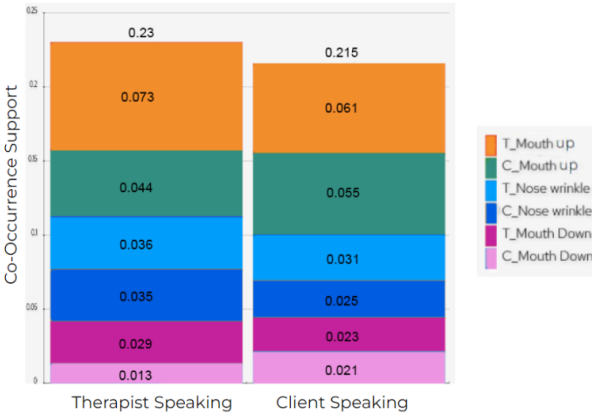

Figure 3: Each cell of the bar chart illustrates the support of co-occurrence between the type of speaker and a facial expression category

In the co-occurrence diagrams, we observe a predominant presence of the $mouth\_up$ categories for both the therapist and the client speaking, as depicted in Figure 3. Notably, the therapist's smile expressions was most prominent when they were speaking. While other facial expressions were also present, their occurrence support values were comparatively lower. A Chi-squared test was built around the matrix as a contingency table. Co-occurrence between each Facial expression category and the type of speaker $p < 0.001$. The consistent presence of smile expression according to the speaker, emphasizes the positive and collaborative nature of Motivational Interviewing sessions.

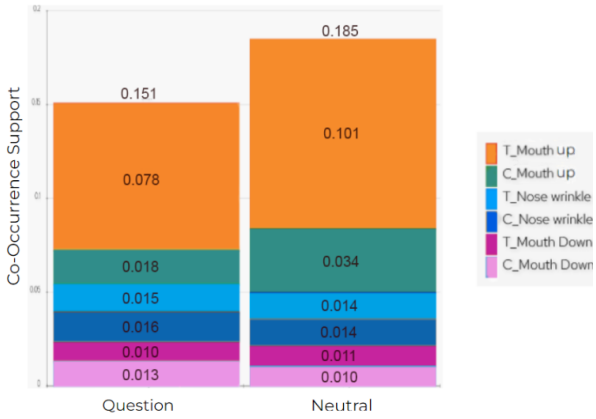

Figure 4: Each cell of the bar chart illustrates the support of co-occurrence between a frequent type of (MI) talk and a facial expression category

To dig deeper into the data, we explored the co-occurrence of facial expression categories in relation to specific MI talk types for both the therapist and client. The type of talks that were most frequent for each interlocutor were $question$ behaviour for therapist, and $neutral$ talk type for the client. The co-occurrence support of these talk types with various facial expression categories are illustrated in Figure 4. Incorporating insights from the diagram, we can see that specific types of talks are associated with a heightened proportion of $mouth\_up$ facial expression categories $p < 0.05$. This trend underscores the inference that certain conversational elements, particularly specific types of talks, elicit a more pronounced activation of positive action units, thereby contributing to the nuanced dynamics of facial expressions within the interaction during counselling sessions

Building on our analysis, we further examined the co-occurrence of facial expression categories with other MI talk types, specifically the $reflection$ behavior from the therapist's perspective and the $change$ talk from the client's side. While the distribution and support of co-occurrence varied compared to the $question$ and $neutral$ talks, the $mouth\_up$ category remained prominently activated. This consistent prominence of the $mouth\_up$ category across different talk types further emphasizes its significance in MI interactions.

this observations directly addresses our research question RQ3. The varied distribution of co-occurrence of the different facial expression categories with varied MI talk types suggests a strong correlation between certain facial expressions and specific MI behaviours. However, when we ventured into analyzing the $sustain$ and $information$ talks, the contingency results between the facial expression categories and these type of talk were not significant. The primary reason for this lack of clarity is the infrequent occurrence of these specific MI talk types in our dataset. Their limited presence meant that they rarely co-occurred in patterns with facial expression categories, making the interpretation of any potential relationships challenging.

5.2. inter speaker action units and Type of talks

In order to investigate role that play social cues in the dynamics of interaction, it becomes necessary to shift our focus to the inter-speaker analysis. This perspective is crucial as it sheds light on the nuanced interplay of facial expressions and talk types between the two speakers, offering insights into the core of interaction. Notably, by examining these facial expressions from an inter-speaker standpoint, we can ascertain that their activation are not merely tied to lip movements during speech (as for $AU26$) but have broader implications in the context of the interaction.

To effectively capture and represent the co-occurrence proportions from an inter-speaker perspective, we adopted the pie chart approach. This choice was influenced by its ability to succinctly depict proportions. The co-occurrence proportion matrix was thus constructed to capture the relative co-occurrence of specific events with others. Specifically, for a given event (e.g., therapist speaks), we first computed the global support of all patterns containing that event. The proportion of co-occurrence between this event and another (e.g., $mouth\_up$ facial expression of the client) was then determined by dividing the support of the pattern containing both events by the global support of the chosen event. Mathematically, the proportion PP of co-occurrence between event A and event B is represented as:

$$P(A,B) = \frac{\text{Support}(A,B)}{\sum \text{Support}(A, \text{all events})} \quad (2)$$

Where:

- $P(A,B)$: Proportion of co-occurrence of events A and B
- Support(A,B) : Support value of the pattern containing events A and B
- Support (A,all events) : Sum of support values of all patterns containing event A

5.2.1. Listener's facial expression category co-occurrences

Building on this methodology, our initial exploration centered on the facial expression categories activated in one listener while the other was speaking, as depicted in Figure 5. This approach not

only provides a clearer understanding of the inter-speaker dynamics but also emphasizes the significance of non-verbal cues in shaping the interaction.

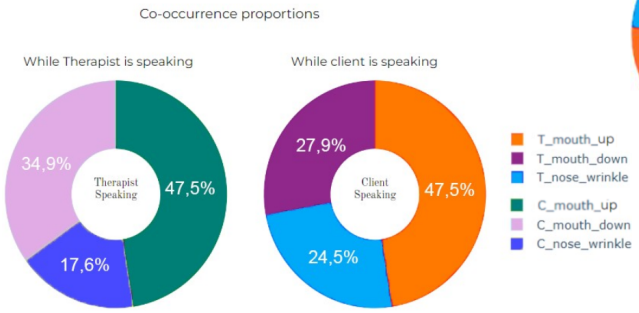

Figure 5: Co-occurrence proportions of listener facial expression catégories

In Figure 5, the prominence of the $mouth\_up$ category is evident for both interlocutors when they are in the listening role, $p < 0.05$. This underscores the significance of the $mouth\_up$ category in the context of the interaction, reinforcing its association with positivity and rapport-building. Additionally, there's a noticeable difference in the $mouth\_down$ category between the two speakers. Specifically, the client exhibits a higher presence of the Mouth Down category when listening compared to the therapist. This observation suggests that while both participants actively engage in non-verbal communication during listening, the therapist places a particular emphasis on conveying positivity. Such findings highlight the pivotal role of non-verbal cues, especially from the therapist's perspective, in shaping and enhancing the quality of the interaction

5.2.2. Facial expression categories dynamics

Subsequently, our attention shifted towards understanding the intricate interplay between facial expression categories. This was driven by our intent to look into behaviours such as mimicry, and adaptive behaviours, all rooted in our theoretical framework. Given our emphasis on the management of positivity within the interaction, we focused on the $mouth\_up$ category, which serves as a hallmark of positive rapport. Specifically, we sought to observe the interactive dynamics between the therapist's and client's $mouth\_up$ activation. We investigated the co-occurrence proportions of the client's facial expression categories when the therapist exhibits a $mouth\_up$ facial expressions and vice versa. The insights from this exploration are presented in Figure 6.

The therapist's inclination towards positive reinforcement is evident. When the client expresses smile through $mouth\_up$ category, the therapist

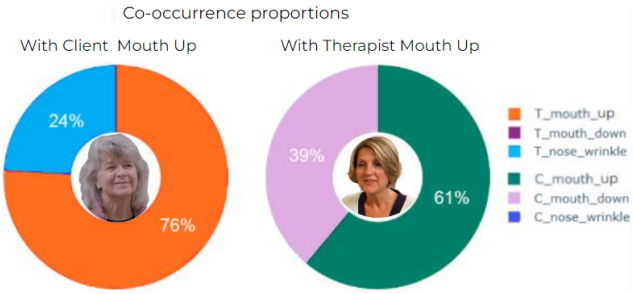

Figure 6: Co-occurrence proportions between therapist and client Mouth up category

frequently reciprocates with the same facial expression category, suggesting a form of mimicry. This behavior emphasizes the therapist's role in fostering a positive rapport and aligning with the client's expression. Conversely, the client, while also showing a tendency to return the therapist's smile expression, does so to a lesser degree. Notably, the $mouth\_down$ category is more pronounced in the client's responses to the therapist's $mouth\_up$. This heightened expression of a potentially negative facial cue in response to positivity may highlight the client's varied expectancy violation and positivity management due to their "passive" role in the context of (MI). The therapist, being the facilitator, is more attuned to maintaining a positive atmosphere, often compensating or reciprocating with positive behaviours to manage and enhance positivity trough the $mouth\_up$ category expression . This dynamic highlights the distinct roles each plays within the therapeutic setting, with the therapist actively working on promoting positivity in someway, while the client navigates passively their expressions. Furthermore, this analysis provides insights into our research question RQ3 The therapist's consistent effort to maintain and reciprocate positive facial cues, especially in response to the client's varied expressions, suggests an adaptive approach aimed at enhancing the quality of the MI session.

This inter-speaker facial expression co-occurrence analysis offers valuable insights into the adaptation of non-verbal behaviours within the context of Motivational Interviewing (MI). Furthermore, it provides a unique perspective on the expectancy violation theory, especially considering the varied valence associated with the facial expression categories we examined.

5.2.3. Positivity decrease

Given the prominence of the $mouth\_up$ categories across various types of talks and the inter speaker analysis phase, we sought to delve deeper into its significance. Considering that the $mouth\_up$ facial expression category is closely

linked to smile and the positivity aspect of rapport in our study, we aimed to validate the theoretical assumption from rapport theory which posits that the importance of positivity diminishes as the interaction progresses (Tickle-Degnen and Rosenthal, 1990).

To investigate this, we plotted the mean occurrence proportion of the $mouth\_up$ category throughout the duration of the interaction. By calculating an average interaction duration based on our dataset, we were able to determine the mean occurrence proportion of the $mouth\_up$ category over time. This allowed us to observe how the significance of positivity, as represented by the $mouth\_up$ categories, evolves during the interaction, particularly concerning the therapist. The findings from this analysis are presented in Figure 7. The results, depicted in the figure, reveal a discernible trend: the activation of $mouth\_up$ categories diminishes as interactions progress. There's a noticeable peak at the onset, which gradually decrease to minimal levels towards the conclusion of the interaction. This trend aligns with the rapport theory's assertion that while positivity is pivotal in the early stages of social interactions, its relevance gradually diminishes. This observation also provides an answer to our research question RQ3. The diminishing prominence of the $mouth_up$ category over the course of the interaction underscores the nuanced role of positivity expressed through therapist's facial expressions, in shaping the interaction quality and non-verbal dynamics of MI sessions.

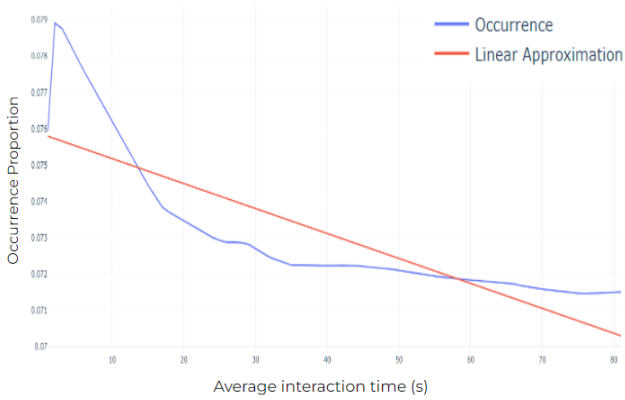

Figure 7: Occurence proportion variation of mouth up category over interaction time

6. Future work and Limitations

Our study provides insights into the impact facial expression categories in MI interactions, but has notable limitations. Firstly, due to dimension reduction issues, we didn't account for the duration and start times of events, which would have enriched our understanding of dynamics like in turn-taking silences. Secondly, our dataset had an imbalance in terms of annotated therapist behaviours and client talk types (like sustain, information and advice), limiting our analysis of less frequent behaviours. Finally, our analysis was based solely on successful and high-quality MI videos. A more comprehensive study would involve contrasting these with unsuccessful MI sessions to truly underscore the positive influence of facial expressions and positivity in the quality of counseling sessions.

Moving forward, our research will address the aforementioned limitations. Leveraging machine learning instead of sequence mining algorithm, we aim to develop an embedding architecture that captures the nuances of facial expression categories in relation to MI's behavior types and talk types while taking into account the timings. Our goal is to design a generative model capable of producing appropriate therapist facial expression categories, adapting dynamically to the progression of counseling sessions and the diverse MI behaviours and type of talks exhibited. Furthermore, testing this model using social robots or virtual agents in both positive and neutral settings will help assess the true impact of facial expressions on session quality.

7. Conclusion

In this research, we addressed the existing gap surrounding the influence of non-verbal behavior in Motivational Interviewing (MI) sessions. Rooted in a theoretical foundation from psychological theories such as the Self Determination Theory, Theory of Rapport, and Expectancy Violation Theory, our study highlights role of non-verbal cues in MI. Through a processing of the AnnoMI database (Wu et al., 2023) and the deployment of sequence mining algorithms (Wang Xin, 2022) , we discerned specific facial expression categories. Our exploration looked into their co-occurrence support in relation to the speakers, the varied MI annotated types of talks, and behaviours, and the intricate dynamics between them. Our observations and statistical analyses enabled us to address our research questions comprehensively. Among our findings, the $mouth\_up$ facial expression category, assimilated to smile expression, emerged as a beacon of positivity management, predominantly exhibited by therapists during MI sessions. This observations not only reaffirms the therapist's active role in fostering a positive therapeutic environment but also paves the way for future studies aiming to further unravel the complexities of non-verbal communication in therapeutic settings. However, our study didn't fully explore the timing and duration of these expressions. Some MI be-

haviours were also underrepresented in our data. As we move forward, we aim to bridge these gaps using cutting-edge machine learning techniques. Our overarching objective remains to refine MI sessions, making them more tailored and impactful for clients. Moreover, our focus on successful MI sessions leaves room for comparison with less successful ones, offering a potential avenue for future research to better evaluate the impact of facial expressions on counseling quality.

8. Acknowledgements

This work was performed as a part of ANR-JST-DFG PANORAMA project.

9. Bibliographical References

Baker, Z. G., Watlington, E. M., and Knee, C. R. (2020). The role of rapport in satisfying one's basic psychological needs. *Motivation and emotion*, 44:329–343.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE.

Berscheid, E. (1994). Interpersonal relationships. *Annual review of psychology*, 45(1):79–129.

Borgelt, C. and Kruse, R. (2002). Induction of association rules: Apriori implementation. In *Compstat: Proceedings in Computational Statistics*, pages 395–400. Springer.

Borsari, B., Hustad, J. T., Mastroleo, N. R., Tevyaw, T. O., Barnett, N. P., Kahler, C. W., Short, E. E., and Monti, P. M. (2012). Addressing alcohol use and problems in mandated college students: a randomized clinical trial using stepped care. *Journal of consulting and clinical psychology*, 80(6):1062.

Burgoon, J. K., Stern, L. A., and Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns.* Cambridge University Press.

Chen, Y.-L. and Hu, Y.-H. (2006). Constraint-based sequential pattern mining: The consideration of recency and compactness. *Decision Support Systems*, 42(2):1203–1215.

Colby, S. M., Orchowski, L., Magill, M., Murphy, J. G., Brazil, L. A., Apodaca, T. R., Kahler, C. W., and Barnett, N. P. (2018). Brief motivational intervention for underage young adult drinkers: Results from a randomized clinical trial. *Alcoholism: clinical and experimental research*, 42(7):1342–1351.

Deci, E. L. and Ryan, R. M. (2012). Self-determination theory. *Handbook of theories of social psychology*, 1(20):416–436.

Dermouche, S. and Pelachaud, C. (2016). Sequence-based multimodal behavior modeling for social agents. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 29–36.

Du, S., Tao, Y., and Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the national academy of sciences*, 111(15):E1454–E1462.

Ekman, P. and Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Klohnen, E. C. and Luo, S. (2003). Interpersonal attraction and personality: What is attractive–self similarity, ideal similarity, complementarity or attachment security? *Journal of personality and social psychology*, 85(4):709.

Klonek, F. E., Wunderlich, E., Spurk, D., and Kauffeld, S. (2016). Career counseling meets motivational interviewing: A sequential analysis of dynamic counselor–client interactions. *Journal of Vocational Behavior*, 94:28–38.

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., and Atkins, D. C. (2015). More than reflections: Empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior therapy*, 46(3):296–303.

Miller, W. R., Moyers, T. B., Ernst, D., and Amrhein, P. (2003). Manual for the motivational interviewing skill code (misc). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*.

Miller, W. R. and Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., and Ernst, D. (2003). The motivational interviewing treatment integrity (miti) code: Version 2.0. *Retrieved from Verfübar unter: www. casaa. unm. edu [01.03. 2005]*.

Nakano, Y. I., Hirose, E., Sakato, T., Okada, S., and Martin, J.-C. (2022). Detecting change talk in motivational interviewing using verbal and facial information. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, pages 5–14.

Nielsen, F. and Nielsen, F. (2016). Hierarchical clustering. *Introduction to HPC with MPI for Data Science*, pages 195–211.

Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., and An, L. (2016). Building a motivational interviewing dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 42–51.

Pérez-Rosas, V., Wu, X., Resnicow, K., and Mihalcea, R. (2019). What makes a good counselor? learning to distinguish between high-quality and low-quality counseling conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 926–935.

Resnicow, K., DiIorio, C., Soet, J. E., Borrelli, B., Hecht, J., and Ernst, D. (2002). Motivational interviewing in health promotion: it sounds like something is changing. *Health Psychology*, 21(5):444.

Rubak, S., Sandbæk, A., Lauritzen, T., and Christensen, B. (2005). Motivational interviewing: a systematic review and meta-analysis. *British journal of general practice*, 55(513):305–312.

Tanana, M., Hallgren, K., Imel, Z., Atkins, D., Smyth, P., and Srikumar, V. (2015). Recursive neural networks for coding therapist and patient behavior in motivational interviewing. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 71–79.

Tickle-Degnen, L. and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293.

Torre, I., Tuncer, S., McDuff, D., and Czerwinski, M. (2021). Exploring the effects of virtual agents' smiles on human-agent interaction: A mixed-methods study. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE.

Tran, T., Yin, Y., Tavabi, L., Delacruz, J., Borsari, B., Woolley, J., Scherer, S., and Soleymani, M. (2023). Multimodal analysis and assessment of therapist empathy in motivational interviews.

Vallerand, R. J. (2000). Deci and ryan's self-determination theory: A view from the hierarchical model of intrinsic and extrinsic motivation. *Psychological inquiry*, 11(4):312–318.

Van Minkelen, P., Gruson, C., Van Hees, P., Willems, M., De Wit, J., Aarts, R., Denissen, J., and Vogt, P. (2020). Using self-determination theory in social robots to increase motivation in l2 word learning. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 369–377.

Wang Xin, Hosseininasab Amin, C. P. K. S. v. H. W.-J. (2022). Seq2pat: Sequence-to-pattern generation for constraint-based sequential pattern mining. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11):12665–12671.

Wu, Z., Balloccu, S., Kumar, V., Helaoui, R., Reforgiato Recupero, D., and Riboni, D. (2023). Creation, analysis and evaluation of annomi, a dataset of expert-annotated counselling dialogues. *Future Internet*, 15(3):110.

Xiao, B., Can, D., Georgiou, P. G., Atkins, D., and Narayanan, S. S. (2012). Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–4. IEEE.