A Prompts

Prompt for generating SOAP note (Figure 7)

```
You are a professional healthcare provider. The following is a conversation between you and a homecare patient.
Please summarize the conversation using the SOAP (Subjective, Objective, Assessment, Plan) note format.

  Input format:
  A transcript of the conversation as a single string.

  Output:
  A JSON object in the following format:{
    "S": "Subjective findings — the patient's reported symptoms, concerns, or experiences.",
    "O": "Objective findings — observable, measurable facts such as vital signs, physical exam findings, or clinician
observations.",
    "A": "Assessment — your clinical judgment or diagnosis based on the subjective and objective information.",
    "P": "Plan — recommended next steps, treatments, referrals, or follow-up care."
  }

  Note: Do not include any other text other than the json object.
  Input:
```

Figure 7: Prompt template used for LLM-based SOAP note summarization from conversation transcripts.

Prompt for LLM-based holistic illness severity scoring (Figure 8)

Prompt for ALM acoustic analysis (Figure 9): We ask the ALM to generate a list of phrases rather than full sentences or fixed categories. This approach is motivated by several considerations: full sentences are difficult to evaluate because the same description can be expressed in many different ways, even though the underlying meaning is the same. Fixed categories, on the other hand, would constrain the model's output.

```
You are a professional healthcare provider reviewing SOAP (Subjective, Objective, Assessment, Plan) notes and
VITAL information from homecare visits. Your task is to evaluate and score the patient's overall health condition.

   Instructions:
   - Lower numbers to indicate better health (e.g., 1 = Good overall health, 5 = Critical condition, hospital referral
recommended.)
   - Provide a brief reason for each ranking.

   Input format:{
      "SOAP_note": {
        "S": "<subjective note>",
        "O": "<objective note>",
        "A": "<assessment>",
        "P": "<plan>"
        },
      "VITAL":{ "BMI": <List of BMI values recorded during the visit>,
                "DIASTOLICBLOODPRESSURE": <List of diastolic blood pressure values recorded during the
visit>,
                "HEIGHT": <List of height measurements recorded during the visit>,
                "PULSE": <List of pulse measurements recorded during the visit>,
                "RESPIRATIONS": <List of respiration rates recorded during the visit>,
                "SYSTOLICBLOODPRESSURE": <List of systolic blood pressure values recorded during the visit>,
                "TEMPERATURE": <List of body temperature readings recorded during the visit>,
                "WEIGHTLBS": <List of body weight in pounds recorded during the visit>,
                "O2SATURATION": <List of oxygen saturation levels recorded during the visit>,
                "BLOODSUGAR": <List of blood sugar levels recorded during the visit>,
                "PAIN": <List of self-reported pain levels recorded during the visit>,
                "WEIGHTKGS": <List of body weight recorded during the visit>
            }
   }

   Note:
   - Each vital sign may be measured multiple times or not at all during a visit.
   - The presence or absence of a measurement may reflect clinical judgment or specific concerns.

   Output:
   A dict object in the following format:{
     "health_condition_score": <score_assigned_to_visit>,
     "reason": "<explanation justifying the assigned score>"
   }

   Note: Do not include any other text other than the dict.
   Input:
```

Figure 8: Prompt template used for LLM-based illness scoring from SOAP notes and vital signs.

You are a specialist in speech and speaker analysis.

You are given an audio sample consisting of several concatenated segments supposedly spoken by the same individual. (but might mixed with other's voice) Your task is to analyze the **acoustic and prosodic characteristics** of the speaker's voice—**not** the content or meaning of the words.
Provide **specific, interpretable observations** as concise phrases or words in the following categories:
- Speaker traits (e.g., perceived age group, gender)
- Prosody
- Fluency
- Articulation
- Emotional state inferred from vocal tone
- Loudness and energy
- Voice quality
- Signs of fatigue, discomfort, or illness: Inferred **solely from voice** if applicable.

Do not transcribe or reference the linguistic content. Analyze only how the speaker sounds.
Return observations as a list of short descriptive phrases

Output format (JSON only):
```json{
"speaker_traits": ["<age>", "<gender>"],
"prosody": ["", "", ...],
"fluency": ["", "", ...],
"articulation":["", "", ...],
"emotion": ["", "", ...],
"loudness_energy": ["", "", ...],
"voice_quality": ["", "", ...]
"signs_of_fatigue_discomfort":[""]
}

Note:
- only output the JSON object. Do not include any additional text, explanation, or formatting.
Input:

Figure 9: Prompt template used for ALM acoustic analysis instructing the model to analyze acoustic characteristics without transcribing linguistic content.

# B Additional Results

## B.1 Summary statistics for LLM-based holistic illness severity scoring (Table 2)

Table 2: Summary statistics for LLM-generated illness scores across different inputs types.

| Model | Input | No ED/HOSP | | | ED/HOSP | | |
|---|---|---|---|---|---|---|---|
| | | Min | Max | Mean | Min | Max | Mean |
| GPT-4.1 | VITAL only | 1.0 | 4.0 | 1.60 | 1.0 | 2.0 | 1.73 |
| | SOAP only | 1.0 | 4.0 | 2.38 | 1.5 | 4.0 | 3.18 |
| | SOAP+VITAL | 1.0 | 4.0 | 2.27 | 1.5 | 4.0 | 3.14 |
| Llama3.1-8B | VITAL only | 3.0 | 5.0 | 4.38 | 3.0 | 5.0 | 4.30 |
| | SOAP only | 1.0 | 5.0 | 2.78 | 1.5 | 5.0 | 3.67 |
| | SOAP+VITAL | 1.0 | 4.5 | 2.67 | 1.5 | 5.0 | 3.39 |
| MediPhi-Instruct | VITAL only | 5.0 | 5.0 | 5.00 | 5.0 | 5.0 | 5.00 |
| | SOAP only | 1.0 | 4.0 | 2.79 | 2.0 | 5.0 | 3.46 |
| | SOAP+VITAL | 1.0 | 5.0 | 3.25 | 2.0 | 5.0 | 3.97 |
| Qwen3-8B | VITAL only | 1.0 | 5.0 | 1.37 | 1.0 | 4.0 | 1.57 |
| | SOAP only | 1.0 | 4.0 | 2.60 | 1.5 | 4.0 | 3.49 |
| | SOAP+VITAL | 1.0 | 4.0 | 2.59 | 1.5 | 5.0 | 3.38 |
```

B.2 Examples of LLM's rationale for illness scores

1 (Good health): The patient demonstrates excellent recovery after the procedure, with normal vital signs, good functional mobility, independence with daily activities and exercise, and no current pain or acute issues. Minor symptoms (occasional dizziness, mild intermittent pain) are self-managed and non-limiting. Discharge from home care is appropriate; following up with specialists is planned for further optimization.

2: Patient is making good post-operative progress with stable vital signs (BP 120/80, pulse 69, respirations 13, temp 98.1). Incision is healing well with no signs of infection, pain is mild and well controlled, and patient is able to perform activities of daily living with minimal assistance. There are no reports of serious symptoms or complications. Continued monitoring and gradual reduction in opioid use is appropriate for this recovery stage. Only mild issues such as minor discomfort and adhesive residue are present.

3: The patient is stable with no acute distress, pain, or new skin issues, and displays improved mobility and motivation. However, there is persistent right-sided paralysis with significant ongoing mobility limitations, elevated blood glucose consistently above target indicating suboptimal diabetes control, and safety concerns regarding stairs. Edema is improving and vital signs are otherwise stable. The condition is chronic and requires ongoing management but does not meet criteria for hospital referral at this time.

4 (Poor health): Patient is elderly, recently hospitalized, and requires ongoing oxygen therapy for chronic respiratory issues with worsening dyspnea. The patient is unable to stand or transfer independently and needs significant physical assistance and equipment for mobility. There is concern for hypoxemia requiring further respiratory management (possible BIPAP), and care is complicated by coordination of multiple services and progressive de-conditioning. Although vitals are stable, the overall functional and respiratory status is poor, indicating high risk but not immediate critical instability.

B.3 Analysis of articulation

Figure 10 and Table 3 present the ALM analysis results for articulation with GPT-4o-audio. In home care visits, articulation is a less informative indicator of health condition, as participants generally speak in their familiar language and therefore articulate their words clearly.

Table 3: Top-3 articulation descriptions for patients per health condition group using TF-IDF analysis. The phrases within each group are ranked from most to least frequently generated by GPT-4o-audio.

Acoustic Feature	**Illness Score**			
	1 (Good health)	**2**	**3**	**4 (Poor health)**
Articulation	Clear Well Pronounced	Clear Well Pronounced	Clear Well Enunciated	Clear Slightly unclear Deliberate

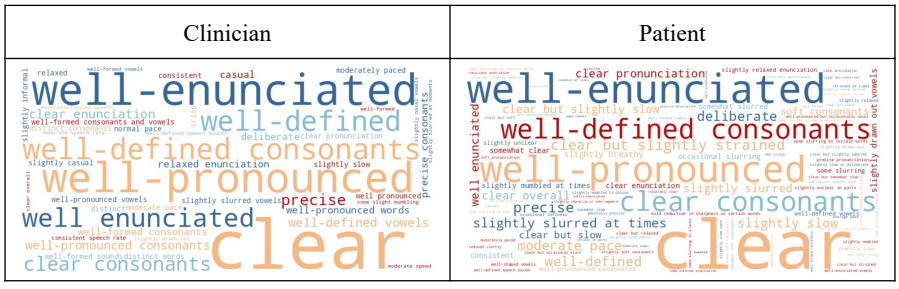

Figure 10: Wordcloud of GPT-4o-audio generated phrase frequencies describing articulation in clinician and patient voices.

C Experimental results of GPT-4o-mini-audio and Qwen2-Audio-7B-Instruct

C.1 Age and Gender Classification with GPT-4o-mini-audio (Figure 11)

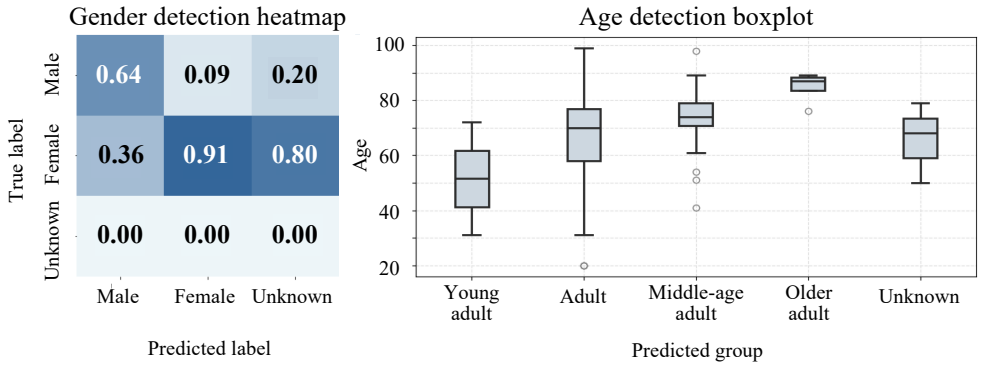

Figure 11: GPT-4o-mini performance in gender and age classification.

C.2 Age and Gender Classification with Qwen2-Audio-7B-Instruct (Figure 12)

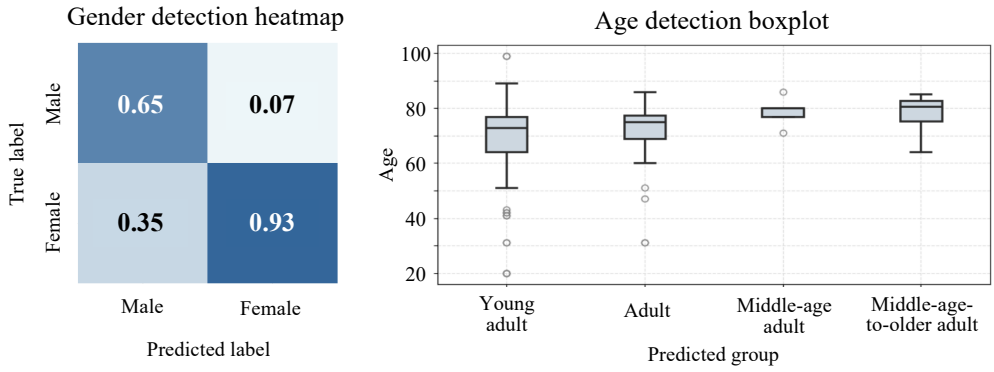

Figure 12: Qwen2-Audio-7B-Instruct performance in gender and age classification.

C.3 Analysis of Clinician and Patient Voices with GPT-4o-mini-audio (Figure 13)

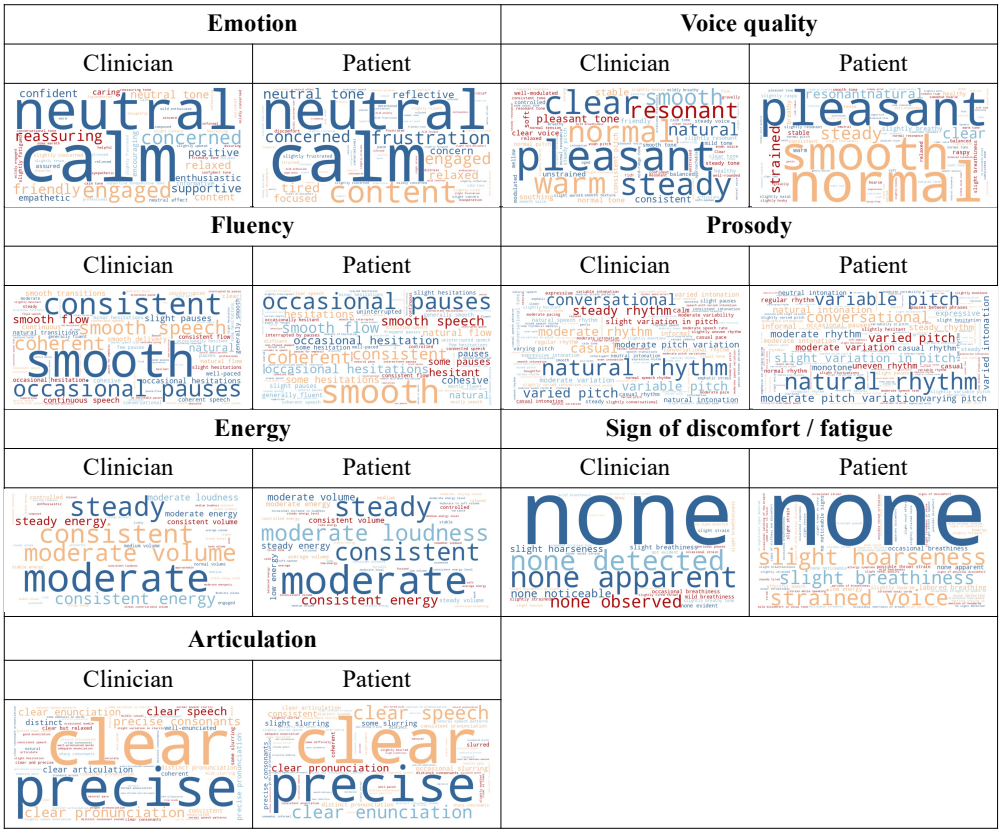

Figure 13: Wordcloud of GPT-4o-mini-audio-generated phrase frequencies for clinician and patient voices.

C.4 Analysis of Clinician and Patient Voices with Qwen2-Audio-7B-Instruct (Figure 14)

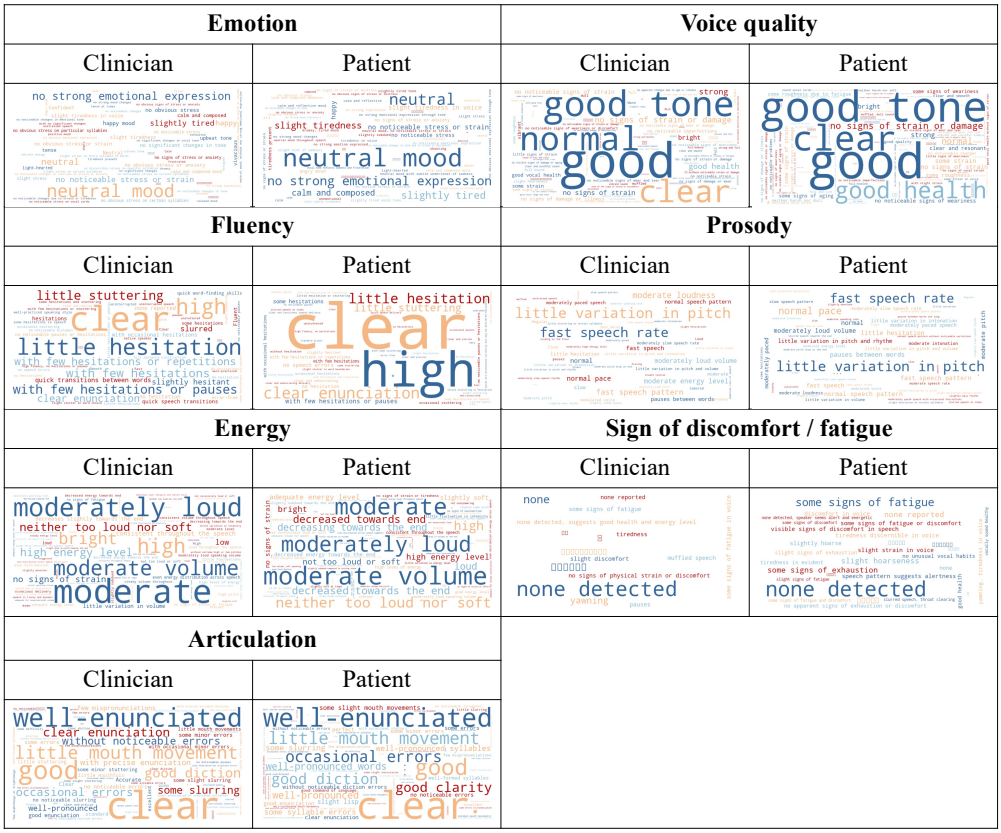

Figure 14: Wordcloud of Qwen2-Audio-7B-Instruct-generated phrase frequencies for clinician and patient voices.

C.5 Analysis of Patient Voices Across Health Conditions with GPT-4o-mini-audio (Table 4)

Table 4: Top-3 GPT-4o-mini-audio-generated acoustic descriptions for patients per health condition group using TF-IDF analysis. The phrases within each group are ranked from most frequent to least frequent.

Acoustic features	**Illness score**			
	1(Good health)	**2**	**3**	**4 (Poor health)**
Emotion	- Calm - Content - Neutral tone	- Calm - Neutral - Engaged	- Calm - Neutral - Frustration	- Calm - Neutral - Tired
Voice quality	- Pleasant - Warm - Slightly breathy	- Smooth - Normal - Pleasant	- Pleasant - Smooth - Strained	- Natural - Slightly raspy - Hoarse
Prosody	- Natural rhythm - Conversational - Even rhythm	- Natural rhythm - Variable pitch - Conversational	- Natural rhythm - Moderate pitch variation - Conversational	- Natural rhythm - Varied pitch - Moderate rhythm
Fluency	- Smooth - Smooth flow - Coherent	- Smooth - Occasional pauses - Smooth speech	- Smooth - Occasional pauses - Hesitations	- Smooth - Occasional pauses - Smooth flow
Articulation	- Clear - Precise - Clear pronunciation	- Clear - Precise - Consistent	- Clear - Precise - Clear enunciation	- Clear - Precise - Slurred
Energy	- Moderate - Consistent energy - Moderate loudness	- Moderate - Consistent - Steady	- Moderate - Steady - Low energy	- Moderate - Consistent volume - Low
Sign of discomfort / fatigue	- None - Mild discomfort in vocal tone - None detected	- None - Slight hoarseness - Occasional breathiness	- None - Slight hoarseness - Labored breathing	- No noticeable signs - Low energy - Strained voice

C.6 Analysis of Patient Voices Across Health Conditions with Qwen2-Audio-7B-Instruct (Table 5)

Table 5: Top-3 Qwen2-Audio-7B-Instruct-generated acoustic descriptions for patients per health condition groups using TF-IDF analysis. The phrases within each group are ranked from most frequent to least frequent.

Acoustic features	**Illness score**			
	1(Good health)	**2**	**3**	**4 (Poor health)**
Emotion	- Neutral mood - No noticeable stress or strain - Calm and reflective	- Neutral mood - Neutral - No strong emotional expression	- Neutral mood - No strong emotional expression - Slight tiredness	- Neutral mood - Neutral - Slightly tired
Voice quality	- Good - Good tone - Normal range of pitch and resonance	- Good - Good tone - Clear	- Good tone - Good - Clear	- Clear - Normal - Some strain
Prosody	- Fast speech rate - Little variation in pitch - Moderately paced speech	- Little variation in pitch - Fast speech rate - Normal pace	- Little variation in pitch - Fast speech rate - Normal pace	- Little variation in pitch - Fast speech rate - Normal
Fluency	- Clear - High - Little hesitation	- Clear - Clear enunciation - High	- Clear - Clear enunciation - High	- Clear - Little hesitation - High
Articulation	- Enunciated - Well - Some slight mouth movements	- Well - Enunciated - Clear	- Well - Enunciated - Clear	- Good - Clear - Clear speech with occasional stuttering
Energy	- Moderate volume - Moderately loud - No signs of fatigue	- Moderate volume - Neither too loud nor soft - Moderate	- Moderate volume - Moderate - Moderately loud	- Moderate - Decreased towards end - Moderately loud
Sign of discomfort / fatigue	- None detected - Slight strain in voice - Throat clearing	- None detected - Some signs of fatigue - No unusual vocal habits	- None detected - Slight hoarseness - Some signs of fatigue	- None detected - Some signs of fatigue - Throat clearing