IBTN 2026 - Abstract

Title: Identifying Multimodal Cues of Ambivalence and Hesitancy for Digital Health Behaviour Change Interventions

Authors: * Manuela González-González[1,2], MSc, Jessica Almeida[1,4], Laura Lucia Ortiz[1], Kim Lavoie[1,4], PhD, Eric Granger[3], PhD, Soufiane Belharbi[3], PhD, Simon L. Bacon[1,2], PhD.

Affiliations:
1. Montreal Behavioural Medicine Centre, CIUSSS du Nord-de-l'Ile-de-Montréal., Montréal, Québec, Canada.
2. Department of Health, Kinesiology, and Applied Physiology, Concordia University, Montréal, Québec, Canada.
3. Department of Systems Engineering, École de technologie supérieure, Montréal, Québec, Canada.
4. Département de Psychologie, Université du Québec à Montréal, Montréal, Québec, Canada.

Background: Sustained behaviour change is challenging and often requires individuals to navigate ambivalence and hesitancy (A&H). While clinicians can detect these states in face-to-face interactions, identifying them in digital contexts remains difficult.
Objective: To characterise multimodal cues associated with A&H in video recordings from adults across Canada, to support the development of machine and deep learning models for automated detection of A&H in real-world digital health interventions.
Methods: A total of 176 participants contributed 718 videos, responding to seven prompts about behaviours toward which they felt neutral, positive, negative, ambivalent, willing, resistant, or hesitant. Three experts in expression recognition annotated videos at four levels: (1) global presence of A&H; (2) frame-level timestamps marking of A&H; (3) modalities involved in A&H (e.g., language, visual); and (4) specific cues indicating A&H (e.g., shoulder shrugging). Descriptive analyses summarised annotation patterns.
Results: Overall, 46.4% of videos showed evidence of A&H, most often in response to prompts designed to elicit these states. Frequent cues included pauses (37.1%) and slowed speech (22.9%) in audio; head shaking (17.4%) and looking away (12.3%) in body language; gaze shifts (25.8%) and eyebrow movements (15.5%) in facial expression; and filler sounds (20.4%) in language. Cross-modal inconsistencies occurred in 37.9% of videos, most commonly between facial expressions and language (42.5%).
Conclusions: A&H are expressed through dynamic, multimodal, and often conflicting cues. These findings provide a foundation for training algorithms to detect A&H patterns and used to improve digital health interventions targeting behaviour change.

Corresponding Author: Manuela González-González, MSc.