Evidence for impaired emotional reactivity in aphasia during naturalistic movie-viewing

Einat Liebenthal
ELIEBENTHAL@MCLEAN.HARVARD.EDU

McLean Hospital, Harvard Medical School https://orcid.org/0000-0003-4917-8160

Manuel Marte
Center for Brain Recovery https://orcid.org/0000-0002-1837-601X

Bryce Gillis
McLean Hospital

Xuehu Wei
Brigham and Women's Hospital,McLean Hospital, Harvard Medical School https://orcid.org/0000-0002-1628-3039

Colin P Galvin
Brigham and Women's Hospital

Laura Rigolo
Brigham and Women's Hospital, Harvard Medical School

Yanmei Tie
Brigham and Women's Hospital, Harvard Medical School

Swathi Kiran

Article

Keywords:

Posted Date: January 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8477032/v1

Additional Declarations: There is **NO** Competing Interest.

Evidence for impaired emotional reactivity in aphasia during naturalistic movie-viewing

Manuel Jose Marte[1,2] Bryce Gillis[2], Xuehu Wei[2,3], Colin Galvin[3], Laura Rigolo[3], Yanmei Tie[3,4], Swathi Kiran[1], & Einat Liebenthal[2,4]

Author Information

Affiliations

Center for Brain Recovery, Boston University[1]
Manuel Jose Marte, Swathi Kiran

Institute for Technology in Psychiatry, McLean Hospital[2]
Manuel Jose Marte, Bryce Gillis, Xuehu Wei, Einat Liebenthal

Department of Neurosurgery, Brigham and Women's Hospital[3]
Xuehu Wei, Colin Galvin, Laura Rigolo, Yanmei Tie

Harvard Medical School[4]
Einat Liebenthal, Yanmei Tie

Contributions

M.J.M. conceptualized the analysis plan, conducted all statistical and machine learning analyses, created visualizations, and wrote the manuscript. S.K. and E.L. conceptualized the original study design and provided supervision. E.L. and Y.T. secured funding. E.L., Y.T., C.G., L.R., X.W., and B.G. developed the movie-viewing paradigm and contributed to methodology refinement. B.G. oversaw and conducted data collection. S.K., E.L., and Y.T. edited the manuscript. All authors reviewed, provided critical feedback, and approved the final manuscript.

Corresponding author

Correspondence to: Einat Liebenthal, eliebenthal@mclean.harvard.edu

Abstract

Language is thought to play a fundamental role in how humans experience emotion. Chronic aphasia, a persistent language impairment, provides a unique window for examining predictions about the role of language in shaping emotions. Here, we tested the hypothesis that aphasia is associated with altered patterns of emotional reactivity during naturalistic movie viewing. We analyzed real-time and holistic emotion ratings, and language comprehension, of movie clips, in persons with aphasia (PWA, n=57) and healthy controls (HC, n=43). PWA versus HC showed reduced typicality and temporal complexity of real-time emotion judgements ($p<.006$), reduced typicality of holistic emotion judgements ($p<0.001$), and reduced language task accuracy ($p<.001$). Principal component analysis yielded a primary component that distinguished the groups ($p<.001$), and in PWA correlated with aphasia ($r=0.644$) and depression ($r=-0.305$) scores. Machine learning achieved strong group classification (AUC=0.964, accuracy=86%) and moderate aphasia severity prediction ($r=0.589$), with both language and emotion contributing features. The findings converge in demonstrating that language and emotion impairments are coupled, during both lexical and discourse level processing. The work highlights that scientific accounts and clinical evaluations of language would benefit from incorporating real-time measures of emotion processing, particularly for improving our understanding of real-world functional communication and impairments.

1. Introduction

Language is thought to play a fundamental role in how humans experience emotion. Contemporary neurobiological accounts suggest that subjective momentary emotional states result from the activation of multiple subcortical and cortical systems, including limbic and paralimbic regions involved in processing core affect and salience, midline regions linked with self-referential processing, and language regions in the inferior frontal and anterior temporal cortex involved in processing emotion concepts[1–2]. Language regions are activated even during emotion induction tasks using non-verbal stimuli, such as viewing emotional films without sound, experiencing self-induced emotional states through imagery, and perceiving facial emotion expressions[1,3]. The distributed neural architecture underlying emotion processing supports constructionist theories proposing that emotions are not biologically hardwired but rather constructed by the brain using language-based conceptual knowledge[4-6]. According to this framework, language helps transform more elemental affective sensations into emotional experiences and memories by providing a conceptual framework for interpretation. Critically, the constructionist framework suggests that language plays a fundamental role in making sense of ongoing sensory perceptions, such that successful emotional responding depends on intact language systems[7]. Thus, chronic aphasia, a persistent speech and language impairment that affects one's ability to communicate, provides a unique window for examining theoretical predictions about the role of language in shaping affective sensations and making emotion judgements. If language is necessary to transform vague affective sensations into defined emotions, then having chronic language impairments should lead to altered patterns of emotional reactivity. Furthermore, emotional reactivity patterns could serve as complementary indicators of language function, particularly during naturalistic tasks that require ongoing conceptual scaffolding.

Despite the theoretical foundation supporting the potential of studying aphasia to illuminate accounts of language-emotion relationships, and the clinical value of identifying emotion deficits in aphasia, surprisingly little research has directly examined emotion processing in this population. Most prior work on emotion in aphasia has focused on using emotional stimuli to improve language performance. A comprehensive review of this work[8] concluded that using emotional (versus neutral) stimuli enhanced performance across several linguistic tasks, though noted limitations included use of few stimuli and stimuli not validated for emotional valence, and tasks that were not ecological. Compared to controlled experimentation which emphasizes a specific task (e.g., identification, discrimination, recall) and uses limited sets of unimodal, static stimuli (e.g., pictures of emotional faces or emotional scenes, emotional written words), naturalistic paradigms such as movie-watching are particularly promising for studying the interaction of language and emotion because they engage perceptual and cognitive processes in parallel, and require real-time integration of information across functional domains to follow the movie plot[9,10]. Movie-watching paradigms also allow for the examination of moment-to-moment dynamics of language and emotion processing, providing insights into the temporal patterns and context effects that characterize real-world communication, which conventional paradigms cannot capture.

In the present study, we tested the hypothesis that chronic aphasia is associated with altered patterns of emotional reactivity, in proportion with the severity of the language impairments. We opted to examine this hypothesis using a naturalistic movie-viewing paradigm with real-time ratings of emotional valence, holistic ratings of specific emotions, and movie comprehension and word retrieval tasks, to gain a better understanding of the contribution of language and emotion deficits to the real-life communication challenges faced by individuals with aphasia. We reasoned that the real-time emotional valence responses may be particularly sensitive to comprehending the unfolding plot, as conveyed by the movie's narrative and audiovisual content. The holistic ratings of specific emotions may be particularly sensitive to comprehending subtle nuances in the meaning of emotion words. Together, these emotion measures may capture subtle comprehension difficulties missed by conventional language tests focused on basic word concepts and single sentences. We therefore analyzed: 1) in the emotion domain, the typicality and temporal complexity of real-time emotional valence ratings, and the typicality of holistic ratings of discrete emotions, and 2) in the language domain, the accuracy of movie comprehension and word retrieval measures, in persons with aphasia (PWA, n=57) relative to healthy controls (HC, n=43), for eight short movie clips. A principal component analysis was applied to understand the structure of the emotion and language responses in relation to group classification and aphasia severity. In addition, the influence of depression and anxiety, which are common co-morbidities in aphasia that can affect language and emotion, was tested in a subset of participants in whom those symptom scores were collected. Finally, to better characterize group differences and clinical relevance, we used a support vector machine (SVM) classifier to identify which movie-based emotional and language features best distinguished PWA from HC, and a support vector regression (SVR) model to examine whether these same features predicted aphasia severity in the PWA group.

2. Results

2.1. Demographic and clinical characterization of the PWA and HC groups

See **Table 1** for the demographic composition and clinical characteristics of each group. The groups differed in age, with PWA being 5 years older on average than HC (PWA: 59.5 ± 10.4 years, HC: 54.3 ± 12.3 years, $t = 2.223$, $p = 0.029$). The groups did not differ in gender distribution (PWA: 42.1% females, HC: 51.2% females) or race (PWA: 21.1% non-white, HC: 16.3% non-white). The subgroups who completed the depression and anxiety tests (PWA: n = 45, HC: n = 33) differed in depression scores (PHQ-8: 5.2 ± 5.7 versus 1.6 ± 2.3, $t = 3.665$, $p < 0.001$) and anxiety scores (GAD-7: 4.7 ± 5.0 versus 0.9 ± 1.3, $t = 4.813$, $p < 0.001$), with PWA reporting mild symptoms and HC no symptoms on average.

2.2. Stimulus set

The stimuli were seven short (140-348 secs) clips from generally well-known live-action movies of different genres selected from the Dynamos database[11], and one from an animated family movie (see **Table 2**). The clips successfully elicited a range of positive and negative emotions, and real-time valence ratings, as designed. In HC, the clip that elicited the highest ratings of positive emotions was from “Akeelah and the Bee” (3.54±1.11 out 5). The clip that elicited the highest ratings of negative emotions was from “No country for old men” (2.37±1.26 out of 5). The real-time emotional valence ratings ranged from a positive mean (across the clip) of 1.78+1.49 for “The parent trap” to a negative mean of -0.72±1.59 for “No country for old men”. These emotion ratings in healthy older adults are consistent with those we observed in healthy younger adults[11].

2.3. Group differences in movie-based language and emotion responses

Language measures

In the language domain, the PWA and HC groups differed on language task accuracy, with PWA scoring lower on both the movie comprehension (HC: 98.9%, PWA: 92.7%; $\beta = -2.080$, SE = 0.459, $z = -4.532$, $p < .001$) and antonym generation (HC: 94.9%, PWA: 59.6%; $\beta = -3.732$, SE = 0.412, $z = -9.060$, $p < .001$) tasks across *clips*. HC performed at near-ceiling levels on both tasks, while PWA showed mild impairment on the movie comprehension task and moderate-to-severe impairment on the antonym generation task, confirming the expected language deficits in the aphasia group. **Figure 1** shows language task scores by group and clip.

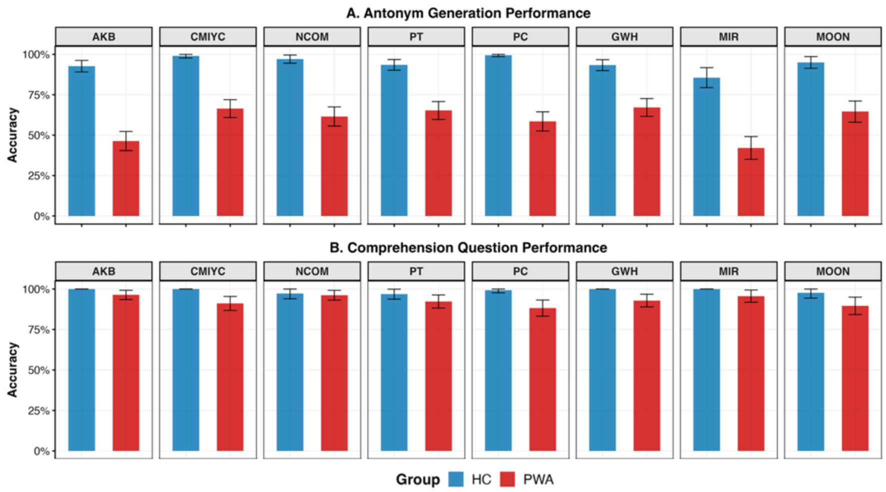

Figure 1. Movie-based language task performance by group and clip. The bar charts display the mean accuracy expressed as a percentage on the Antonym Generation (Panel A) and Comprehension Question (Panel B) tasks, for each of the eight movie clips (movie acronyms listed in Table 2), in HC (blue bars) and PWA (red bars). Error bars represent 95% confidence intervals (standard error of the mean).

Emotion measures

In the emotion domain, the individual ratings were evaluated with respect to the normative baseline, which was computed for each feature as the mean of the HC group. The distributions of the holistic ratings of positive and negative emotions are shown for all clips in **Figure 2**. For the holistic ratings, the degree of deviation (measured as mean squared z-score - MSZ) from the HC group reference was significantly higher in PWA than HC (estimate = 0.529, SE = 0.148, t = 3.578, $p < .001$)

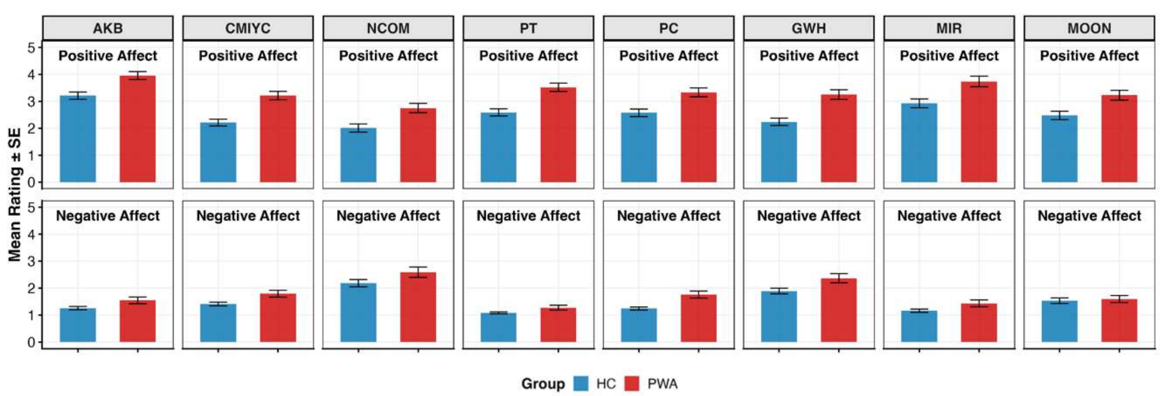

Figure 2. Mean holistic ratings of positive and negative emotions for each movie clip in PWA and HC. The bar charts display mean Positive and Negative Affect Schedule Short-Form (PANAS-SF[12]) ratings aggregated across the five positive (top panels) and five negative (bottom panels) items, respectively, for each clip, in HC (blue) and PWA (red). Error bars represent 95% confidence intervals (standard error of the mean). Ratings for each item were made on a discrete scale from 1 ("very slightly or not at all") to 5 ("extremely").

The trajectories of the real-time emotional valence ratings are shown for all clips in the two groups in **Figure 3**. There were no group differences in the degree of deviation (measured as MSZ) of the real-time emotional valence ratings from the HC group reference (estimate = 0.303, SE = 0.202, t = 1.502, p = .136), suggesting that both groups were overall able to track the emotional content of the clips. However, there were significant group differences in the temporal pattern and complexity of the real-time valence ratings. The PWA ratings were significantly less synchronized (measured as time-resolved intersubject correlation - ISC) with the HC group reference than the HC ratings (estimate = -0.110, SE = 0.039, t = -2.833, p = .006). The PWA ratings also showed significantly lower entropy than the HC ratings, specifically at the medium (interaction coefficient = -0.018, SE = 0.005, t = -3.584, p < .001) and long (interaction coefficient = -0.013, SE = 0.006, t = -2.339, p = .019), but not the short, time scales. This finding suggests that PWA exhibited more predictable, less complex emotional response patterns when integrating information over longer temporal windows (20-35 seconds). For entropy, age also had a significant effect, with older participants showing lower entropy across time scales (estimate = 0.0018, SE = 0.0009, t = 2.00, p = .048).

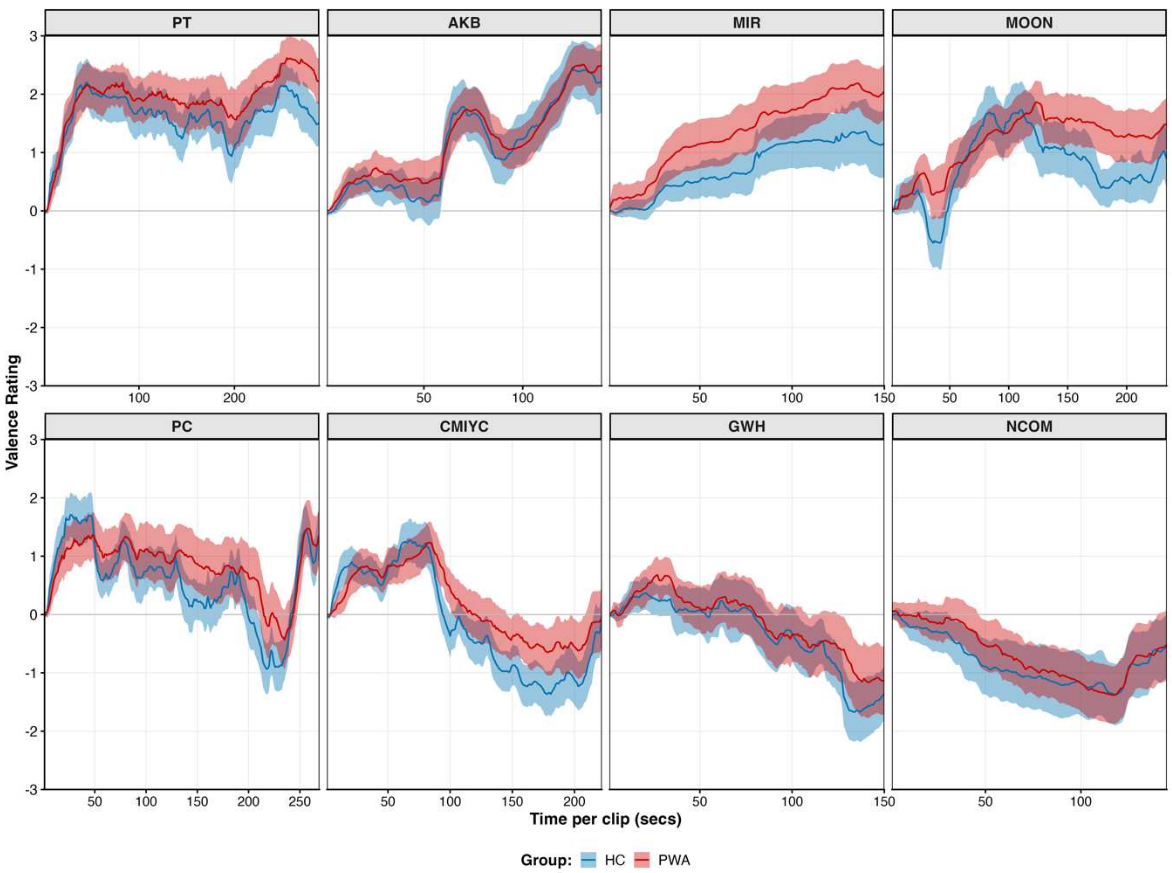

Figure 3. Real-time emotional valence ratings during movie-watching in persons with aphasia and healthy controls. Mean real-time emotional valence ratings (solid lines) and 95% confidence intervals (shaded regions) in PWA (red) and HC (blue) for each clip (movie acronyms listed in Table 2), sorted from the most positive to the most negative clip on average. Ratings were made on a continuous scale from -4 (most negative) to +4 (most positive), with 0 indicating neutral valence. Error bands represent 95% confidence intervals (standard error of the mean) across subjects in each group.

2.4. Principal component analysis of language and emotion movie responses

A principal component analysis that included the language accuracy scores, holistic MSZ deviation scores, and real-time MSZ, ISC and entropy scores, from both groups, resulted in three components (PC1-PC3) with eigenvalues ≥1, accounting for 66.9% of the total variance in the movie responses (PC1: 30.7%, PC2: 19.6%, PC3: 16.6%, see **Figure 4**). Bootstrap cross-validation with 1,000 iterations confirmed the stability of this component structure: PC1 explained 32.3% of the variance (95% CI: 26.7%–40.3%, CV = 0.106), PC2 explained 20.4% (95% CI: 17.8%–23.5%, CV = 0.072), and PC3 explained 15.9% (95% CI: 13.4%–18.2%, CV = 0.076). We further tested for group differences in each of the components. PC1 scores differed reliably between groups (HC-PWA difference = 1.618, SE = 0.222, $t = 7.28$, $p < 0.001$; 99.8% of bootstrap samples significant at $p < 0.01$, mean bootstrap $p = 0.0004$). In contrast, PC2 scores showed no group difference (HC-PWA difference = -0.297, SE = 0.228, $t = -1.31$, $p = 0.195$; only 21.9% significant at $p < 0.01$, mean bootstrap $p = 0.292$). PC3 scores differed between groups but with low reliability (HC-PWA difference = 0.588, SE = 0.217, $t = 2.71$, $p = 0.008$; only 32.5% significant at $p < 0.01$, mean bootstrap $p = 0.192$). Because PC1 captured more of the variance in the movie language and emotion scores than PC2 and PC3, and was the only component that reliably differed between PWA and HC, the subsequent analyses and discussion focus on this component.

PC1 captured the structure of both the language tasks and the emotional reactivity measures across the groups. PC1 explained 31.0% of the variance in movie comprehension, 29.9% in antonym generation, 20.8% in holistic emotion ratings, and 9.7% in real-time emotional valence ratings. Higher PC1 scores represented higher response accuracy on the two movie language tasks and more typical emotional reactions, measured as lower deviation of the holistic and real-time emotion ratings, and higher synchrony of the real-time ratings.

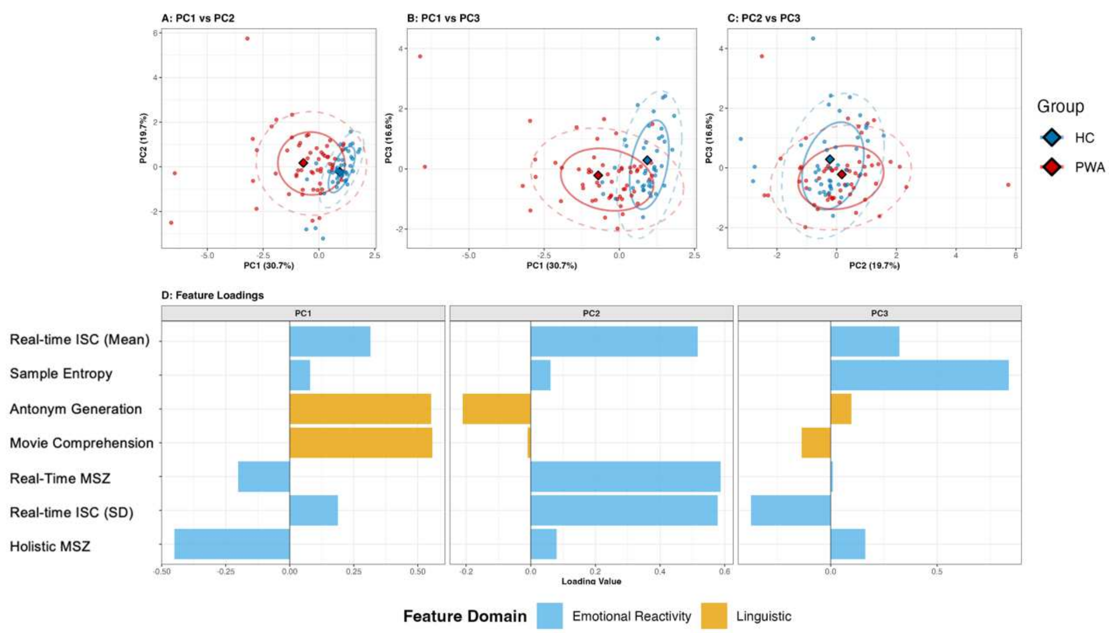

Figure 4. Distribution of language and emotion principal component scores in PWA and HC. (A-C) Scatter plots showing the distribution of scores for principal components in PWA (red) and HC (blue): (A) PC1 vs PC2, (B) PC1 vs PC3, and (C) PC2 vs PC3. The diamonds in each plot indicate the PWA (red) and HC (blue) group centroids. The ellipses show 68% and 95% confidence intervals around each centroid. (D) Feature loadings for PC1-3, showing the contribution of emotional reactivity (light blue) and linguistic (orange) features to each principal component.

2.5. Correlations between PC1 and clinical scores

Within the aphasia group, PC1 scores were positively correlated with the aphasia severity scores (AQ: r = 0.644, p < 0.001). PC1 was also negatively correlated with the depression scores (PHQ-8: r = -0.305, p = 0.041), but not the anxiety scores (GAD-7: r = -0.240, p = 0.117). In addition, the depression and anxiety scores were not significantly correlated with the aphasia severity scores (PHQ-8: r = -0.117, p = 0.443; GAD-7: r = -0.105, p = 0.499). Within the healthy control group, no significant correlations were observed between PC1 and the depression and anxiety scores (p>0.24). Thus, in the aphasia group, better performance on the movie language and emotion tasks was associated with milder clinical aphasia, and independently with milder depression scores (see **Figure 5**).

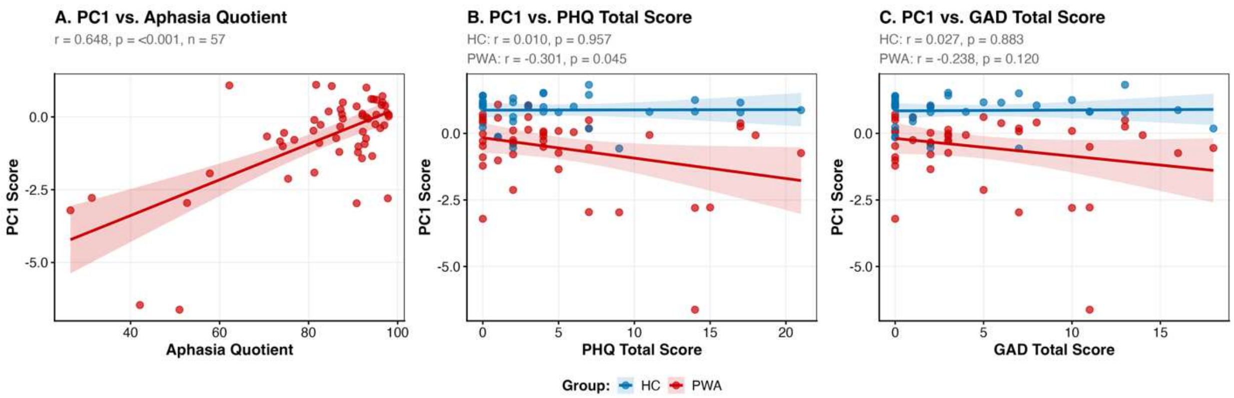

Figure 5. Correlations between principal component 1 (PC1) and clinical measures. Scatter plots and correlation lines showing the relationship between PC1 and (A) the aphasia severity quotient for the PWA (red) group; and (B) the PHQ-8 depression and (C) GAD-7 anxiety scores for both the PWA and HC (blue) groups. Solid lines represent the linear model fit for each correlation. Error bands (shaded areas) represent 95% confidence intervals (standard error of the mean) across subjects in each group.

2.6. Machine learning modeling

Group classification

The support vector machine classifier based on the movie language and emotion response features successfully discriminated PWA from HC with high accuracy (see **Table 3** for model performance statistics, **Figure 6** for visualization). Cross-validation yielded a mean AUC of 0.964 (95% CI: 0.916–0.985), with overall accuracy of 86.0% (95% CI: 78.1–91.4%).

The SHAP analysis revealed that the linguistic features contributed meaningfully to group classification, with antonym generation ranking first (mean |SHAP| = 0.205), and movie comprehension (mean |SHAP| = 0.059) ranking third, in importance. Emotional reactivity measures also contributed meaningfully, with deviation in holistic ratings ranking second (mean |SHAP| = 0.099), intersubject correlation in real-time ratings ranking fourth (mean |SHAP| = 0.045), and entropy of real-time ratings ranking fifth (mean |SHAP| = 0.035).

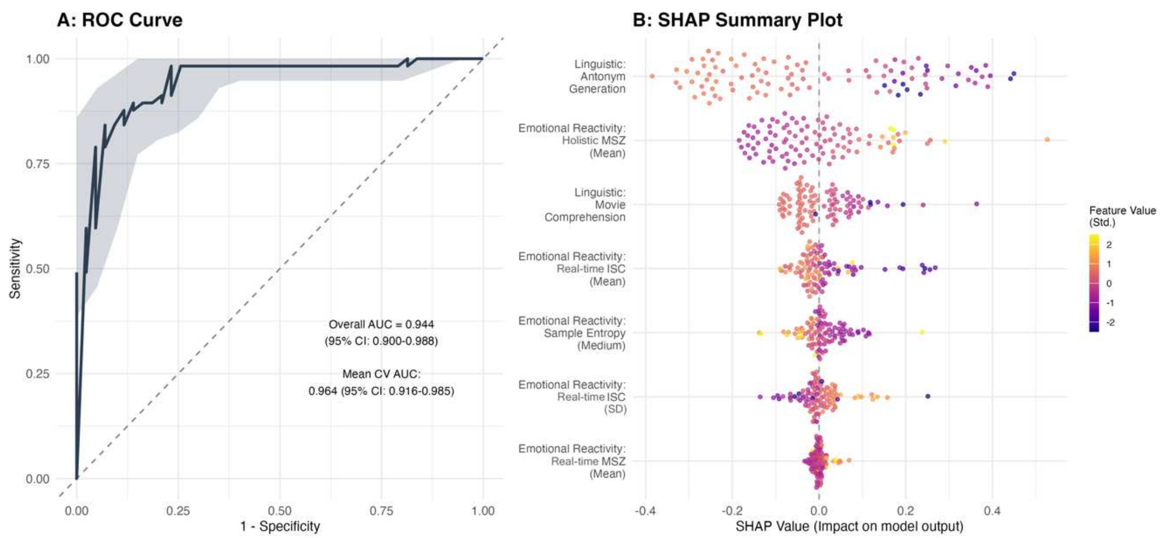

Figure 6. Machine learning classification of PWA versus HC based on movie-viewing performance features. (A) Receiver operating characteristic (ROC) curve showing classification performance of support vector machines with radial basis function kernels. (B) SHAP (SHapley Additive exPlanations) values representing feature contributions to model predictions. The features are ranked by their mean absolute SHAP value. Each point represents one participant-fold instance, with color indicating standardized feature value (yellow = high, purple = low). Positive SHAP values (right side) push the model toward classifying a participant as HC, while negative SHAP values (left side) push toward classifying a participant as PWA. For most features, lower positive or more negative SHAP values are associated with lower movie performance and higher likelihood of classification as PWA (and the opposite is true for MSZ deviation and ISC variability).

Prediction of aphasia severity

Within the aphasia group, the support vector regression model predicted the aphasia quotient scores with moderate accuracy (**Figure 7**). Leave-one-out cross-validation yielded a root mean square error (RMSE) of 13.22 points (95% CI: 10.71–17.48) and mean absolute error (MAE) of 8.92 points (95% CI: 7.01–12.25), indicating typical prediction errors of approximately 7–12 AQ points. There was a significant correlation between the actual and model-predicted scores (95% CI: 0.036–0.733; R = 0.589). SHAP analysis revealed that antonym generation was the dominant predictor of aphasia severity (mean |SHAP| = 4.99), contributing substantially more than all other features. The remaining features showed similar, modest contributions (mean |SHAP| = 1.90 or less).

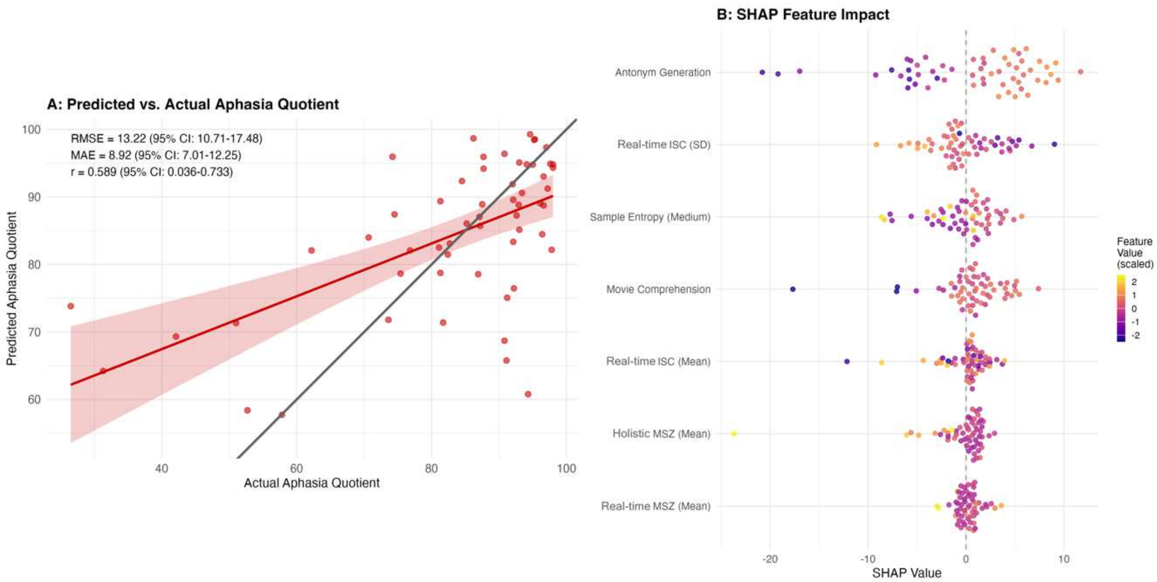

Figure 7. Support vector regression analysis for predicting aphasia quotient scores. (A) Predicted versus actual Aphasia Quotient (AQ) scores. The solid red line indicates the linear model fit and the shaded area represents the 95% confidence interval. The gray line represents perfect prediction. (B) SHAP summary plot showing the contribution of each feature to the model's output. Positive SHAP values push predictions toward higher AQ scores (less severe aphasia) and negative SHAP values push predictions toward lower AQ scores (more severe aphasia). Points are colored by the standardized feature value (high = yellow, low = purple).

3. Discussion

In this work, we tested the hypothesis that individuals with aphasia have altered patterns of emotional reactivity during naturalistic movie viewing, and that the extent of emotional disruption is associated with aphasia severity. To this end, we conducted a comprehensive study of real-time emotional valence ratings, holistic ratings of specific emotions, and language comprehension responses, to short movie clips, in PWA and HC. The findings converge in demonstrating associated emotion and language deficits in aphasia. PWA relative to HC showed reduced typicality and temporal complexity of the real-time valence judgements ($p<.006$), reduced typicality of the holistic emotion judgements ($p<0.001$), and reduced language task accuracy ($p<.001$). Principal component analysis yielded a primary component that captured 30.7% of the variance in both language and emotion movie responses, distinguished between the two subject groups ($p<.001$), and in PWA correlated with both the aphasia ($r=0.644$) and depression ($r=-0.305$) severity scores. Support vector machine learning achieved excellent group classification (AUC=0.964, accuracy=86%) and moderate aphasia severity prediction ($r=0.589$), based on both language and emotion discriminating and predictive features. Overall, the findings show that impairments in language and emotion are coupled, highlighting that emotion should be considered in neuroscientific accounts and clinical assessments of language.

The findings of emotion deficits in aphasia add to a small body of prior work, showing that impairment in semantic access following left-hemisphere stroke is associated with deficits in facial emotion recognition[13]. Deficits in face emotion recognition were also observed across several variants of primary progressive aphasia, suggesting that different language deficits can contribute to the emotional impairment[14]. However, to our knowledge the relationships between language and emotion impairments have not previously been investigated during naturalistic behaviors and at the scale of the present study. Our naturalistic movie-viewing paradigm captured the dynamic, moment-to-moment interplay between language comprehension and emotional experience during complex, multimodal narrative processing, as well as the post-viewing emotional reactions. We found a primary component (PC1, **Figure 4**) that explained significant variance in the language scores (~30%), holistic emotion ratings (21%), and real-time emotion ratings (10%), suggesting that these linguistic and emotion measures of movie processing are coupled. The finding that both the holistic and real-time emotion ratings load on PC1 and are abnormal in aphasia further suggested that language and emotion interact at multiple levels. Language may provide the conceptual framework for labeling and making holistic judgements about specific emotions experienced during movie watching (e.g., excited, inspired, upset, scared), as suggested for basic facial expressions in prior work[13-14]. Language, however, also enables viewers to construct appropriate emotional valence judgements (positive, negative, neutral) in real-time based on comprehension of the unfolding narrative, characters' motivations, contextual cues, etc. The strong correlation between PC1 and the aphasia severity quotient observed here suggests that in individuals whose language systems are more compromised, this conceptual architecture deteriorates, resulting in holistic and real-time emotional responses that are more idiosyncratic and deviate from normative patterns.

The machine learning analyses provided convergent validation of PC1's constituent features as clinically meaningful markers of aphasia. The support vector machine classifier achieved excellent discrimination between the groups, with SHAP analysis confirming that the features loading heavily on PC1 were indeed the most important for classification, including (in descending order of importance) antonym generation, holistic rating of specific emotions, movie comprehension, and real-time emotional valence rating. The convergence between unsupervised dimensionality reduction and supervised classification strengthens the interpretation that interactive language-emotion function represents a meaningful dimension of cognitive organization that is disrupted in aphasia.

Within the aphasia group, support vector regression successfully predicted aphasia severity, with antonym generation emerging as the dominant predictor. This finding is unsurprising given that the antonym generation task requires both comprehension of the movie dialogue and ability to access and retrieve semantically related words, two core language processes that are commonly disrupted in aphasia[15]. However, the prominence of real-time emotional valence features as the next best predictors suggests that the naturalistic emotion measures may capture functional impairments beyond those captured by standard aphasia batteries, and which therefore carry potential for enhancing clinical assessment. The finding that the trajectory and temporal complexity of the real-time emotional valence ratings were predictive of aphasia severity suggests that dynamic processing and integration of contextual information in real-time is also disrupted in aphasia. Such deficits in real-time processing are expected to impact discourse comprehension and conversational abilities.

Overall, the findings carry important clinical and scientific implications for our understanding of aphasia. A main outcome is that clinical evaluation would benefit from extending beyond traditional language batteries to include measures of emotional function, and measures of real-time discourse processing, particularly for assessing real-world functional impairments. Recent consensus statements have emphasized the need for more functional communication assessments that capture real-world communication challenges[16]. At the lexical level, our findings with the movie paradigm demonstrate that using word descriptors to rate specific emotions captures variance beyond that captured by antonym generation. The narrative-level measures (comprehension questions, real-time emotional valence judgement) further explained variance beyond that captured by the lexical measures. The more comprehensive characterization of language and emotion deficits in aphasia afforded by the naturalistic movie-viewing assessment could provide a basis for devising targeted interventions. For example, an individual in our sample with mild language issues based on clinical assessment (AQ=92), showed highly atypical movie emotion ratings consistent with significant difficulties in discourse comprehension, suggesting that they would benefit from interventions focused on functional communication, rather than lexical fluency.

The observed relationships between movie responses in the language and emotion domains, with both aphasia and depression severity, further reinforce the potential value of assessing emotional function in aphasia, and shed new light on depression in aphasia. Depression and anxiety are known to affect 30–70% of individuals with aphasia[17,18]. Our findings show that while language and emotion deficits were associated with aphasia and depression severity, the aphasia and depression severity scores themselves were not significantly correlated with one another. This pattern suggests that factors beyond language deficits contribute to depression in aphasia, for example the emotion deficits could exacerbate depression independently.

From a theoretical standpoint, the emergence of a primary dimension encompassing both language and emotion performance is consistent with psychological constructionist theories proposing that language plays an active role in emotional processing[5-7]. The fact that the primary component explained more variance in both emotion and language domains, and showed the strongest correlations with clinical measures of language and emotion deficits, suggests that language and emotion processing are tightly coupled during naturalistic narrative comprehension.

Several limitations warrant consideration in interpreting these findings. First, while the study was well powered for detecting group differences and correlations with clinical scores, subgroup differences between aphasia types could not be investigated. The sample was predominantly composed of individuals with anomic aphasia, which is the most common aphasia type but may not be representative of the full spectrum of language-emotion interactions across aphasia profiles. Future work should investigate whether the relationships between language and emotion processing differ systematically across aphasia types, particularly in individuals with primarily expressive versus receptive or mixed deficits. Similarly, investigations of larger samples with varying levels of aphasia and depression severity would serve to disentangle independent, interactive, and causal effects of language and emotion deficits on functional communication outcomes. Second, the present study included only a single-word production task, which may have underestimated the role of emotion-language interaction during connected speech production. The brief spoken narrative summaries of each clip that were collected but not analyzed here, could in future work shed more light on language-emotion interactions during expression versus comprehension tasks, and their respective contributions to functional communication deficits in aphasia. Finally, the absence of functional neuroimaging measures precludes a complete understanding of the neural underpinnings of the behavioral dimension representing interactive language and emotion processing that we identified in this study. Future studies combining our paradigm with functional neuroimaging in aphasia could clarify the neural networks participating in various levels of language-emotion interactions, and in turn inform both theoretical models of language and emotion processing and targeted neuromodulation or network-based rehabilitation approaches.

In summary, this study is the first large-scale investigation using naturalistic movie-viewing to examine language and emotion interactions in aphasia. Our novel approach of combining real-time tracking of emotional valence during movie-viewing with post-movie discrete emotion and language tasks revealed a key finding: the interaction of language and emotion emerged as a primary dimension distinguishing persons with aphasia from healthy controls. This dimension was correlated with aphasia severity, and independently with depression severity. Unsupervised dimensionality reduction and supervised machine learning provided converging evidence that language-emotion interaction is a core function that is systematically disrupted in aphasia. These findings shed light on the challenges faced by individual with aphasia during real-world communication. From a clinical perspective, our findings challenge the traditional separation of language and emotion in aphasia assessment and treatment. As we move toward more personalized diagnostic and rehabilitation strategies, understanding how language and emotion interact during naturalistic tasks may help identify treatment targets and optimize outcomes for persons living with aphasia.

4. Methods

4.1. Participants and Clinical Characterization

Participants in the study were 58 adults with stroke-induced aphasia and 60 healthy controls, ages 30-80 years. One participant in the PWA group was excluded due to difficulty performing the paradigm within the allotted time, and 17 participants were excluded from the analyses of the HC group because they were classified as having mild cognitive impairment based on Montreal Cognitive Assessment scores <26 (MoCA[19]). Thus, 57 PWA and 43 HC are included in this report.

PWA were recruited from the Boston University Center for Brain Recovery database using informed consent procedures approved by the Boston University Charles River Campus Institutional Review Board and the Mass General Brigham (MGB) Institutional Review Board. Inclusion criteria consisted of a single left-hemisphere stroke at least 6 months prior to testing, aphasia diagnosis confirmed by medical records, Western Aphasia Battery-Revised[20] (WAB-R) Aphasia Quotient (AQ) of 25 or greater, premorbid right-handedness, native or native-like English fluency, adequate (or corrected to adequate) vision and hearing as needed for task completion, and ability to provide informed consent independently or with appropriate support. Exclusion criteria included history of other neurological conditions, including extensive cognitive or physical disabilities that would prevent performing the paradigm's tasks. HC participants were recruited from the greater Boston community through Rally, an MGB-mandated research participation platform, following informed written consent procedures approved by the MGB Institutional Review Board. Eligible healthy volunteers met the same inclusion/exclusion requirements as the aphasia group but had no history of stroke, neurological disorder, psychiatric illness, or speech-language impairment. In addition, HC had to score 26 or greater on the MoCA, for their data to be included in the analyses.

A subset of participants (45 PWA and 33 HC) who enrolled after the protocol was amended in this regard, were additionally assessed for depression and anxiety using the Patient Health Questionnaire-8[21] (PHQ-8) and the Generalized Anxiety Disorder-7[22] (GAD-7). The effects of depression and anxiety on PC1 and AQ were assessed only in this subsample.

4.2. Movie Clip Stimuli

The stimuli were eight short clips (140–348 seconds) from generally well-known live-action movies (e.g., "Good Will Hunting"), that are freely available in open libraries. Seven of the clips were selected based on having a simple dialogue and narrative, and eliciting reliable real-time emotional valence judgements (positive and negative) in neurotypical adults, as validated in the Dynamic Affective Movie Clip Database for Subjectivity Analysis[11] (DynAMoS). Several of these clips were trimmed (relative to their length in DynAMoS) to reduce overall testing time while preserving the core emotional and narrative arc. The eighth clip was taken from "Partly Cloudy", an animated movie with no dialogue, to examine emotional processing independent of dialogue comprehension. The clips were extracted from high-definition source material and presented at 1920×1080 resolution, with audio levels normalized across clips to ensure consistent presentation quality. Subtitles were not displayed to avoid potential confounding effects of different reading abilities. The stimuli characteristics and emotional response profiles in the HC group are presented in Table 2.

4.3. Experimental Procedures

PWA completed two sessions. The first was a one-hour remote session for comprehensive neuropsychological testing, consisting of the WAB-R to determine the Aphasia Quotient, the Sentence Comprehension subtest of the Northwestern Assessment of Verbs and Sentences[23] (NAVS), the Elevator Strings subtest from the Test of Everyday Attention[24] (TEA) to assess sustained auditory attention, the Trail Making Test Parts A and B[25] (TMT) to evaluate processing speed and executive function, the PHQ-8 for depression screening, and the GAD-7 for anxiety screening. The second two-hour session was devoted to the movie-watching paradigm and conducted on a separate day, either in person if the participant was able to come to the Lab, or remotely if they were not. The remote sessions were conducted via Zoom video conferencing in a quiet room at the participant's residence, and the in-person sessions were conducted in a testing room dedicated to movie-viewing experiments at McLean Hospital. The HC participants all completed a single in-person, two-hour session starting with brief cognitive screening (MoCA), followed by the movie-watching paradigm.

In the movie-viewing session, participants were first trained to use the Continuous Affect Rating and Media Annotation (CARMA) software[26] for continuous rating of emotional valence on a scale from "Very Negative" (-4) to "Very Positive" (+4). Participants controlled the slider using a joystick in the in-person setting (which all HC and 20 of the 57 PWA participated in) or arrow keys in the remote Zoom setting (which 37 of 57 PWA participated in), to indicate how each moment of the movie made them feel. Training was conducted with a practice clip not included in the experimental set, and proceeded until participants demonstrated an understanding and correct use of the slider to rate emotional valence in real-time. Input method (joystick vs. keyboard) did not significantly influence any of the real-time emotional reactivity measures (linear mixed-effects models with input method as a fixed effect, controlling for aphasia severity and age, with random intercepts for participant and clip, yielded a chance probability of .549 or greater).

In the experiment, participants viewed as many of the 8 clips as they could complete within the two-hour session, presented in randomized order. Of 848 possible participant-clip combinations, 744 (87.7%) were successfully completed. Before playing each clip, the experimenter provided a standard one-sentence neutral description for context (e.g. "This clip features a conversation between two men in a restaurant" for the clip from *Catch Me If You Can.* See **Table 2** for all descriptions). Following each clip, participants were asked whether they had been able to hear the audio and see the video clearly, and whether they had previously seen the clip. They then indicated the extent of specific emotions elicited by each clip on a discrete scale from 1 ("very slightly or not at all") to 5 ("extremely"), using the Short-form Positive and Negative Affect Schedule[12] (PANAS-SF). Finally, they completed three language tasks: 1) A 60-second free narrative summary of the clip content using the prompt "Tell me in your own words what happened in the clip"; 2) Three two-alternative forced-choice comprehension questions designed to assess basic understanding of plot elements, character relationships, visual elements (e.g., locations) and key narrative events; and 3) Five antonym generation trials using word prompts drawn from the clip's dialogue or visual scenery (e.g., "beautiful", as Akeelah's final word 'pulchritude' is defined to mean "beautiful" by the judge). Participants had 4 seconds to generate an antonym for each word.

4.5. Language and Emotion Features

Performance on the comprehension and antonym generation tasks was computed for each individual and clip as the ratio of correct responses to total items presented, yielding accuracy scores ranging from 0 to 1. The free narrative summaries of each clip were not analyzed here.

The real-time emotional valence ratings were acquired with CARMA at 30Hz and preprocessed using custom R scripts. The rating time series were down sampled to 1Hz, and the first 10 seconds of responses to each clip were excluded from analysis, as they were considered an orientation period. Three features were computed for the individual real-time responses to each clip: (i) sample entropy, (ii) intersubject correlation, and (iii) mean squared z-score deviation.

Sample entropy - a measure of predictability of future values based on past patterns, representing the temporal complexity and internal structure of time series - was computed using a non-overlapping sliding window at three temporal scales, short (window sizes: 5, 10, 15 seconds), medium (window sizes: 20, 25 seconds), and long (window sizes: 30, 35 seconds). Other parameters included an embedding dimension of $m = 2$ and tolerance threshold $r = 0.15$ times the standard deviation of each time series, following established practices in physiological signal analysis[27]. Lower entropy values indicate more predictable, potentially more rigid response patterns, while higher values suggest more complex, less predictable response patterns.

Time-resolved intersubject correlation (ISC) was applied here to measure the similarity between individual emotional trajectories and group reference patterns. ISC was computed using a 15-second sliding window and 5-second step size. For each window, the Pearson correlation between an individual's ratings and the reference ratings was calculated and Fisher z-transformed to normalize the correlation values. These Fisher-transformed correlation values were then averaged across windows to produce a single ISC value per participant per clip[28,29]. For PWA, the reference trajectories corresponded to the mean ratings across all healthy control participants. For HC, a leave-one-out approach was used where each participant's trajectory was compared to the mean of all other healthy controls, to prevent artificial inflation of correlations. Windows with zero variance were excluded from analysis.

Mean squared z-score (MSZ) deviation was computed to capture overall extent of divergence from normative response patterns. For each participant-clip combination, individual CARMA ratings at each timepoint were z-scored relative to the HC group mean and standard deviation at that timepoint, squared, and then averaged across all timepoints to produce an MSZ value. Higher MSZ values indicate greater deviation from typical response patterns. As with ISC, a leave-one-out approach was implemented for HC participants.

MSZ was also computed for the holistic PANAS ratings, to quantify deviation from normative patterns of discrete emotion recognition. For each of the 10 PANAS-SF items (5 positive; 5 negative), individual ratings were z-scored relative to the mean and standard deviation across all participants for that clip-item combination. The squared z-scores were then averaged across all 10 items within each participant-clip to produce a single MSZ value. This metric captured how atypical an individual's overall pattern of discrete emotion ratings was compared to normative responses for each clip.

4.6. Statistical Analysis

Linear mixed-effects models were used to assess the effects of group membership (PWA, HC) on the real-time emotional reactivity measures, including age and clip as covariates. Age was included in the models because the groups differed in age, and age has been shown to interact with emotion processing[30,31]. Clip was included as a random effect to account for variability across the eight different movie clips: DV ~ Group + Age + (1|Participant) + (1|Clip). For sample entropy, models also included Window Size (short, medium, long) and its interaction with Group. For language tasks scored as binary (1 = correct, 0 = incorrect), generalized linear mixed-effects models with binomial family and logit link were used: Accuracy ~ Group + Age + (1|Stimulus) + (1|Clip) + (1|Participant). Within-PWA analyses examined relationships between the movie measures and aphasia severity (WAB-R AQ) using Pearson correlations. FDR correction for multiple comparisons was applied to these exploratory correlations, and all reported p-values reflect this correction.

4.7. Principal Component Analysis

To select a parsimonious set of features for subsequent multivariate analyses, we conducted pair-wise correlations between the 12 computed movie response features, and selected a subset of 7 that were representative of the language and emotion domains, and for which correlation values were below 0.7. The final feature set included, for the real-time emotional valence measures: 1) Sample entropy (medium scale, selected as representative of the three scales), 2) Deviation from the HC mean (computed as MSZ), and 3-4) Correlation with the HC mean trajectory (computed as mean and standard deviation of ISC, to capture both the degree and consistency of alignment with normative responses, respectively). For the holistic PANAS ratings: 5) Deviation from the HC mean (computed as MSZ). For language: 6-7) Accuracy on the antonym generation and movie comprehension tasks, respectively.

Principal component analysis (PCA) was performed to identify latent dimensions underlying these seven features. The features were standardized prior to PCA using the prcomp function in R. Components were retained based on the Kaiser criterion[32] (eigenvalues ≥ 1.0). The stability of the component structure was assessed using bootstrap validation with 1,000 iterations, and 95% confidence intervals and coefficients of variation (CV = SD/Mean) to estimate the variance explained by each component across iterations.

PC scores were compared between PWA and HC using linear models, with age as a covariate. The models took the form: PC ~ Group + Age, implemented using base R statistical functions (i.e., lm()). Within the PWA group, we examined correlations between PC scores and the WAB-R Aphasia Quotient (AQ) using Pearson correlations. In the subset of participants with available data, we also examined correlations between PC scores and the depression (PHQ-8), and anxiety (GAD-7), scores. To assess the robustness of significant correlations, we conducted outlier detection using multiple converging criteria (Cook's distance, studentized residuals, leverage) and performed sensitivity analyses excluding influential observations.

4.8. Support Vector Machine Classification

To assess group (PWA, HC) classification based on the seven-feature set, we implemented support vector machines (SVM) with radial basis function kernels using the e1071 package. Nested cross-validation prevented overfitting, with 5 outer folds for performance evaluation and 3 inner folds for hyperparameter optimization (cost C = {0.1, 1, 10}; gamma γ = {0.01, 0.1, 1}). Feature standardization occurred within each fold to prevent data leakage. Model performance metrics included accuracy, sensitivity, specificity, precision, F1 and area under the receiver operating characteristic curve (AUC). Feature importance was assessed using SHapley Additive exPlanations (SHAP) values via the kernelshap package[33].

4.9. Support Vector Regression for Aphasia Severity Prediction

To evaluate whether the same features could predict clinical severity (expressed as WAB-R Aphasia Quotient scores) within the PWA group, we implemented support vector regression (SVR) with radial basis function kernels. We used nested leave-one-out cross-validation, with the inner loop optimizing hyperparameters (cost C = {0.1, 1, 10}; epsilon ε = {0.01, 0.1}), while the outer loop evaluated model performance. Model performance was assessed using root mean square error (RMSE), mean absolute error

(MAE), Pearson correlation coefficient, and R^2. Feature importance was again quantified using SHAP values.

All analyses were conducted using R version 4.2.3 with significance set at $p < .05$. Degrees of freedom for mixed-effects models were approximated using Satterthwaite's method[34] in the lmerTest package.

5. Data Availability

The DynAMoS stimulus and emotional response database can be found here: https://dynamos.gitlab-pages.partners.org/dynamos/. Request to use the DynAMoS database can be submitted via the website. Requests to obtain de-identified data from this study, including continuous emotional valence ratings, discrete emotion ratings, language task performance, and neuropsychological assessments, will be considered in accordance with study protocols and ethical approvals, and with respect to participant confidentiality protections and institutional review board requirements. Requests should be directed to the corresponding author (E.L.).

6. Code Availability

All custom R code used for statistical analyses, data preprocessing, machine learning modeling, and figure generation will be made available upon publication. Code includes implementations of principal component analysis, linear mixed-effects models, support vector machines, support vector regression, SHAP value calculations, and bootstrap validation procedures. Code and detailed documentation will be deposited in a public repository <Github URL to repository be provided upon acceptance>.

7. References

1. Saarimäki, H. *et al.* Discrete Neural Signatures of Basic Emotions. *Cerebral Cortex* **26**, 2563–2573 (2016).
2. Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. The brain basis of emotion: A meta-analytic review. *Behav Brain Sci* **35**, 121–143 (2012).
3. Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G. & Just, M. A. Identifying Emotions on the Basis of Neural Activation. *PLOS ONE* **8**, e66032 (2013).
4. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* **12**, 1–23 (2017).
5. Brooks, J. A. *et al.* The role of language in the experience and perception of emotion: a neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience* **12**, 169–183 (2017).
6. Satpute, A. B. & Lindquist, K. A. At the Neural Intersection Between Language and Emotion. *Affec Sci* **2**, 207–220 (2021).
7. Lindquist, K. A., MacCormack, J. K. & Shablack, H. The role of language in emotion: predictions from psychological constructionism. *Front. Psychol.* **6**, (2015).
8. Schwen, B. D. & Harnish, S. M. A Scoping Review on the Effects of Emotional Stimuli on Language Processing in People With Aphasia. *Journal of Speech, Language, and Hearing Research* **65**, 4327–4345 (2022).
9. Hasson, U. *et al.* Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research* **2**, 220–231 (2009).
10. Kauttonen, J., Hlushchuk, Y. & Tikka, P. Optimizing methods for linking cinematic features to fMRI data. *NeuroImage* **110**, 136–148 (2015).
11. Girard, J. M., Tie, Y. & Liebenthal, E. DynAMoS: The Dynamic Affective Movie Clip Database for Subjectivity Analysis. *Int Conf Affect Comput Intell Interact Workshops* **2023**, (2023).
12. Mackinnon, A. *et al.* A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences* **27**, 405–416 (1999).
13. Souter, N. E., Lindquist, K. A. & Jefferies, E. Impaired emotion perception and categorization in semantic aphasia. *Neuropsychologia* **162**, 108052 (2021).
14. Marshall CR, Hardy CJD, Russell LL, Bond RL, Sivasathiaseelan H, Greaves C, Moore KM, Agustus JL, van Leeuwen JEP, Wastling SJ, Rohrer JD, Kilner JM, Warren JD. The functional neuroanatomy of emotion processing in frontotemporal dementias. *Brain* **142**, 2873–2887 (2019).
15. Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R. & Schwartz, M. F. Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cogn Affect Behav Neurosci* **19**, 1286–1298 (2019).
16. Wallace, S. J. *et al.* A core outcome set for aphasia treatment research: The ROMA consensus statement. *Int J Stroke* **14**, 180–185 (2019).
17. Edelkraut, L. *et al.* Spectrum of neuropsychiatric symptoms in chronic post-stroke aphasia. *World J Psychiatry* **12**, 450–469 (2022).
18. Hilari, K. The impact of stroke: are people with aphasia different to those without? *Disability and Rehabilitation* **33**, 211–218 (2011).
19. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **53**, 695–699 (2005).
20. Kertesz, A. Western Aphasia Battery--Revised. https://doi.org/10.1037/t15168-000 (2012).
21. Kroenke, K. *et al.* The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders* **114**, 163–173 (2009).

22. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* **166**, 1092–1097 (2006).
23. Thompson, C. K. Northwestern Assessment of Verbs and Sentences (NAVS). https://nulive.technologypublisher.com/tech/Northwestern_Assessment_of_Verbs_and_Sentences_(NAVS) (2012).
24. Robertson, I., Ward, T., Ridgeway, V., Nimmo-Smith, I. & Anespie, C. The Test of Everyday Attention. Manual. (1994).
25. Reitan, R. M. Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills* **8**, 271–276 (1958).
26. Girard, J. M. CARMA: Software for continuous affect rating and media annotation. *J Open Res Softw* **2**, e5 (2014).
27. Costa, M., Goldberger, A. L. & Peng, C.-K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **71**, 021906 (2005).
28. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).
29. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci* **14**, 667–685 (2019).
30. Gray, V., Douglas, K. M. & Porter, R. J. Emotion processing in depression and anxiety disorders in older adults: systematic review. *BJPsych Open* **7**, e7 (2021).
31. Ruffman, T., Henry, J. D., Livingstone, V. & Phillips, L. H. A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews* **32**, 863–881 (2008).
32. Kaiser, H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* **20**, 141–151 (1960).
33. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. Preprint at https://doi.org/10.48550/arXiv.1705.07874 (2017).
34. Satterthwaite, F. E. An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin* **2**, 110–114 (1946).

8. Acknowledgements

We thank Tyler K. Perrachione for critical comments on the manuscript. We are deeply grateful to all participants with aphasia and healthy control participants, as well as their caregivers and families, for their generous time, effort, and contributions to this research.

9. Ethics declarations

The authors declare no competing interests.

Table 1. Demographic and clinical characteristics of the aphasia and control groups. Differences in group means are noted, using t-tests for continuous variables, chi-square test for gender and race (white vs. non-white), and Fisher's exact test for ethnicity (Hispanic/Latino vs. not Hispanic/Latino). PWA = Persons with aphasia; HC = Healthy controls; WAB-R = Western Aphasia Battery-Revised; MoCA = Montreal Cognitive Assessment; NAVS = Northwestern Assessment of Verbs and Sentences; TEA = Test of Everyday Attention; TMT = Trail Making Test (Part B time / Part A time); PHQ-8 = Patient Health Questionnaire-8; GAD-7 = Generalized Anxiety Disorder-8.

Characteristic	Persons with Aphasia	Healthy Controls	Group Differences
Sample size, n	57	43	—
Demographic Characteristics			
Age (years), mean + SD	59.5 ± 10.4	54.3 ± 12.3	$t = 2.223$, $p = 0.029$
Range	38–80	30–69	—
Gender, n (%)			$\chi^2 = 0.486$, $p = 0.486$
Female	24 (42.1%)	22 (51.2%)	—
Male	33 (57.9%)	21 (48.8%)	—
Race/Ethnicity, n (%)			$\chi^2 = 0.001$, $p = 0.981$
Asian	5 (8.8%)	3 (7%)	—
Black/African American	6 (10.5%)	4 (9.3%)	—
Other/Not specified	1 (1.8%)	0 (0%)	—
White	45 (78.9%)	36 (83.7%)	—
Ethnicity, n (%)			Fisher's exact $p = 1$
Hispanic/Latino	2 (3.5%)	1 (2.3%)	—
Not Hispanic/Latino	54 (94.7%)	42 (97.7%)	—
Unknown	1 (1.8%)	0 (0%)	—
Clinical Characteristics			
Time post aphasia onset (years), mean + SD	7.6 ± 7.3	—	—
Range	0.5–45.9	—	—
WAB-R Aphasia Quotient, mean + SD	83.8 ± 16.4	—	—
Range	26.5–98	—	—
MoCA Score, mean + SD	—	28.2 ± 1.4	—
Range	—	26–30	—
NAVS Sentence Comprehension (%), mean + SD	82.6% ± 17.8%	—	—
Range	40–100%	—	—
TEA Elevator Strings (scaled), mean + SD	6.4 ± 1.3	—	—
Range	1–7	—	—
TMT-B/A Ratio, mean +SD	3.24 ± 1.65	—	—
Range	1.15–8.78	—	—
PHQ-8 Total, mean + SD	5.2 ± 5.7	1.6 ± 2.3	$t = 3.665$, $p < 0.001$
Range	0–21	0–8	—
GAD-7 Total, mean + SD	4.7 ± 5.0	0.9 ± 1.3	$t = 4.813$, $p < 0.001$
Range	0–18	0–4	—

Table 2. Movie stimulus characteristics and mean emotional responses in healthy adults. Basic description of the eight movie clips used in the study, and mean and standard deviation of the holistic ratings of specific positive and negative emotions, and real-time emotional valence ratings, in HC and PWA. The one sentence neutral description of each clip was read to participants before testing that clip. The positive and negative specific emotions were assessed using the Short-form Positive and Negative Affect Schedule (PANAS-SF) on a discrete scale from 1 to 5. The real-time emotional valence ratings were collected on a continuous scale from -4 to +4.

Movie Title	Genre	Year	Director	Clip duration (secs)	Description	Positive Emotions Mean (SD)	Negative Emotions Mean (SD)	Real-time Valence Mean (SD)
Akeelah and the Bee (AKB)	Drama	2006	Doug Atchison	140	*This clip features a contest between two school children.*	3.54 (1.11)	1.39 (0.63)	1.14 (1.48)
Catch Me If You Can (CMIYC)	Crime / Drama	2002	Steven Spielberg	223	*This clip features a conversation between two men in a restaurant.*	2.65 (1.09)	1.58 (0.68)	0.04 (1.49)
Good Will Hunting (GWH)	Drama	1997	Gus Van Sant	150	*This clip features a conversation between two men in an office.*	2.69 (1.18)	2.10 (1.08)	-0.22 (1.67)
Miracle (MIR)	Sports / Drama	2004	Gavin O'Connor	150	*This clip features a scene about a sports team.*	3.29 (1.23)	1.28 (0.57)	1.12 (1.34)
Moonlight (MOON)	Drama	2016	Barry Jenkins	234	*This clip features a conversation between a man and a boy.*	2.83 (1.11)	1.56 (0.73)	1.01 (1.37)
No Country for Old Men (NCOM)	Thriller / Crime	2007	Coen Brothers	145	*This clip features a conversation between two men in a convenience store.*	2.36 (1.01)	2.37 (1.26)	-0.72 (1.59)
Partly Cloudy (PC)	Animation / Short	2009	Peter Sohn	348	*This clip features scenes about a cloud and an animal.*	2.90 (1.15)	1.47 (0.74)	0.71 (1.62)
The Parent Trap (PT)	Family / Comedy	1998	Nancy Meyers	290	*This clip features scenes between a man and a girl.*	3.00 (1.11)	1.17 (0.51)	1.78 (1.49)

Table 3. Support vector machine classification performance metrics. Means and 95% confidence intervals derived from nested cross-validation of support vector machine classification of group (PWA, HC) based on movie language and emotion response features (5 outer folds for performance evaluation, 3 inner folds for hyperparameter optimization). AUC = area under the receiver operating characteristic curve; ROC = receiver operating characteristic; PWA = Persons with aphasia; HC = Healthy controls.

Metric	Mean	95% CI	Description
AUC	0.964	(0.916 – 0.985)	Area Under the ROC Curve
Accuracy	0.860	(0.781 – 0.914)	Overall classification accuracy
Sensitivity	0.845	(0.662 – 0.939)	True positive rate for PWA
Specificity	0.886	(0.593 – 0.977)	True negative rate for HC
Precision	0.845	(0.662 – 0.939)	Positive predictive value for PWA
F1 Score	0.873	(0.798 – 0.923)	Harmonic mean of precision and sensitivity