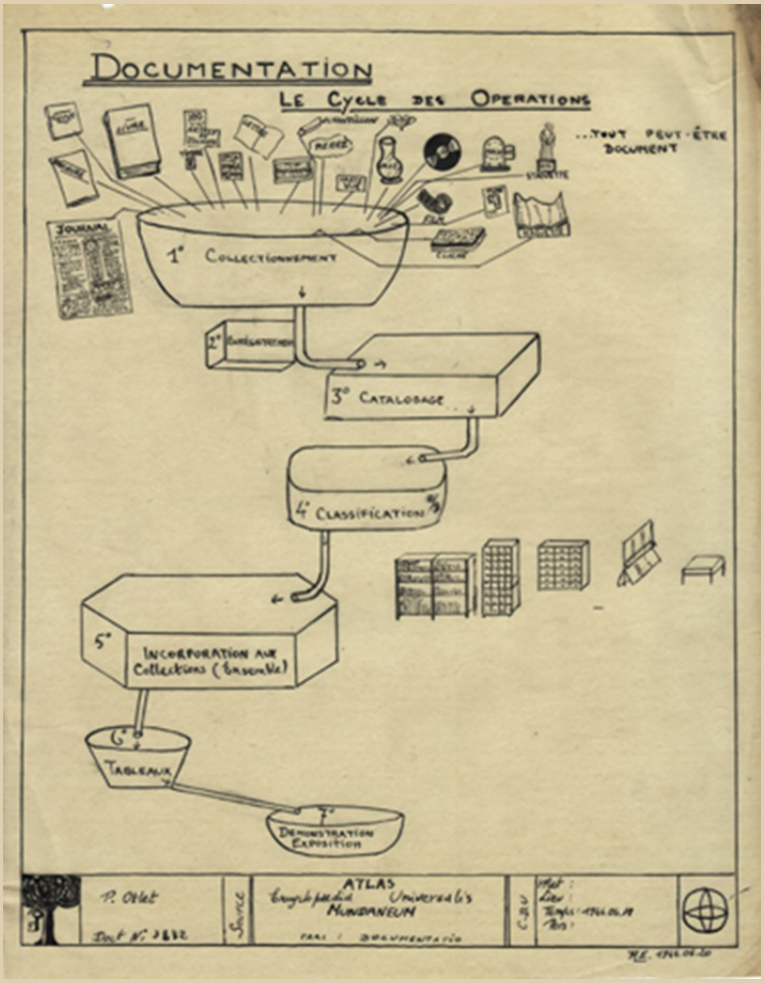
AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 1-2
ANNO 42
GENNAIO-GIUGNO 2024



CACUCCI EDITORE
BARI

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:
Università della Calabria

Direttore Scientifico:
Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:
Fabrizia Flavia Sernia

Editrice: Cacucci Editore S.a.s.
Via D. Nicolai, 39 – 70122 Bari (BA)
www.cacucciditore.it
e-mail: riviste@cacucciditore.it
Telefono 080/5214220

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti le Scienze dell'Informazione, la Documentazione, la Gestione Documentale e l'Organizzazione della Conoscenza. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi possono riguardare topics quali Documentazione, Scienze dell'informazione e della comunicazione, Scienze del testo e del documento, Organizzazione e Gestione della conoscenza, Terminologia, Statistica testuale e Linguistica computazionale e possono illustrare studi sperimentali in domini specialistici, casi di studio, aspetti e risultati metodologici conseguiti in attività di ricerca applicata, presentazioni dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Settore 11/A4 e censita per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali, così come dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) che la annovera tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE –Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; Letteratura Professionale Italiana – Associazione Italiana Biblioteche; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Summon™ – by SerialsSolutions; Ulrich's; UniCat – Union Catalogue of Belgian Libraries; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti dal comitato scientifico sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 42

N. 1-2 – gennaio-giugno 2024

CACUCCI EDITORE

BARI

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
http://www.cacucciedit ore.it e-mail: info@cacucci.it

Sommario

Editoriale

Contributo su invito

Contributi

Testimonianze

Contributi

AIDAinformazioni
ISSN 1121–0095
ISBN 979-12-5965-307-6
DOI 10.57574/596540766
pag. 117-144 (gennaio-giugno 2024)

Discovering and documenting brilliance

A novel multimodal annotation method

Alexander Murzaku*, Pontish Yeramyan, Curt Anderson, Steven Buxbaum, Ruben Diaz, Marielle Lerner, Armenui Minasyan, Hazel Mitchley, Jodie-Ann Pennant, Mia Shang, Brisa Speier Brito

Abstract: The pursuit of understanding human brilliance has long fascinated scholars, practitioners, and observers across various domains of performance. Capturing and understanding human brilliance is an important endeavor, as it allows us to gain insight into the motivations behind extraordinary achievements. However, brilliance remains a complex phenomenon that eludes simple characterization. In this paper we propose a methodology for investigating brilliance. We use structured interviews with high-achieving individuals to document the language surrounding their accomplishments, and code this language using seven distinct themes. Each interview is then annotated using a multimodal annotation schema that captures a variety of linguistic and paralinguistic features, including phonetic information, hand gesture, and eye gaze. This system allows us to discover and document the tacit knowledge that underlies human brilliance and make inroads in understanding the full communicative expression of brilliance.

Keywords: Multimodal Annotation, Tacit Knowledge, Thematic Analysis, Corpus Analysis, Human Brilliance.

1. Introduction

The pursuit of understanding human brilliance has long fascinated scholars, practitioners, and observers across various domains of performance, be it in the arts, sciences, sports, or leadership. Efforts to define brilliance, as well as the exploration into what drives exceptional performance, follow various approaches. Simonton uses his historiometric approach to reverse engineer genius, by investigating «the most eminent figures of history, whether creators or leaders, using various at-a-distance quantitative methods, including computer content analysis of primary documents such as creative products

* Saint Elizabeth University, Morristown, NJ, USA and Leveraging Genius Institute at Gap International, Springfield, PA, USA. amurzaku@steu.edu and alex.murzaku@gapinternational.com. All co-authors are members of the Leveraging Genius Institute at Gap International and can be contacted at first.last@gapinternational.com.

or private correspondence» (Simonton 2016, 4). However, brilliance remains a complex phenomenon that eludes simple characterization. This complexity necessitates innovative methodologies to uncover the underlying elements that compose it.

The methodology described in this paper explores what contributes to brilliance through interviews and offers the possibility of a nuanced analysis facilitated by multimodal annotations. When participants articulate their journeys, philosophies, and reflections, they provide clues to the mental frameworks and attitudes that underpin their performance. However, the richness of these verbal accounts often contains subtleties and nuanced expressions of thought and emotion that traditional analysis methods may overlook. Multimodal annotation is a method that leverages multiple data types — textual, auditory, visual, and kinesthetic cues — to enrich the interpretation of conversation. This composite approach allows researchers to not just listen to what is said but to also tune into how it is expressed. Multimodal analysis offers a glimpse into not only what people do during their expression of brilliance and how they do it, but also what drives them, what triggers their moments of genius. This paper proposes a method to capture and document what creates brilliance.

2. Brilliance as tacit knowledge

Human history is filled with many documented examples of brilliant minds and talents that have influenced the path of human progress. From Socrates to Shakespeare, from Da Vinci to Einstein, from Beethoven to John Lennon, each has left an indelible mark on their respective fields, demonstrating how individual brilliance can drive forward entire domains of human endeavor.

The role of brilliance in human evolution can be seen as a catalyst for growth, for adaptation, and for transformation. It has enabled humanity to transition from basic survival to advancing human knowledge and technology, pushing the boundaries of what it means to be human and how we understand and interact with the world around us.

Examples of brilliance are manifested in the corporate world as well through innovative business strategies, groundbreaking products that we all use, effective leadership, and transformative organizational cultures. Brilliant minds like Steve Jobs at Apple, Jeff Bezos at Amazon, Larry Page and Sergey Brin at Google, or Elon Musk at Tesla show that brilliance does not rest solely on individual genius but also on the ability to foster innovation, adapt to change, and execute visionary strategies that drive businesses — and sometimes entire industries — forward.

While brilliance often involves a unique combination of talent, creativity, and insight that seems inherently personal, organizations can implement practices to capture and disseminate these brilliant sparks. Capturing and

documenting the insights, methodologies, and creative processes of brilliant individuals in knowledge repositories allows organizations to preserve this valuable information. However, it remains in the domain of information; what is missing is capturing brilliance in action, brilliance in performance.

Brilliance in performance, or genius, refers to the outstanding display of skill, talent, and ability routinely transcending standard expectations. It's the combination of innate aptitude, rigorously honed skills, deep understanding, and often an innovative or creative approach that separates merely good performances from truly exceptional ones.

Documenting and preserving brilliance in performance, whether it pertains to artistic achievements, stellar scientific contributions, or exceptional leadership, involves capturing not only the tangible outcomes of such efforts but also the processes, insights, and nuances that characterize them. This is a multifaceted endeavor that demands careful attention to both the tangible and intangible aspects of achievements. What is collected in this endeavor are two forms of knowledge labeled as explicit knowledge and tacit knowledge (Polanyi 1958). Brilliance lives in tacit knowledge.

Explicit knowledge is verbalized, documented and easily transferred. In contrast, tacit knowing is a way of knowing without telling, highlighting the nature of certain types of knowledge - knowledge that one can act upon but finds difficult to fully articulate or transfer solely through written or spoken word (Polanyi 1966, 18). «[T]acit knowledge is knowledge hidden from the consciousness of the knower» point out Olomolaiye and Egbu (2005, 770), and Casonato and Harris (1999, quoted in Dampney, Busch, and Richards 2002, 3) observe that «[t]acit knowledge includes skills, experiences, insight, intuition and judgment. It is typically shared through discussion, stories, analogies and person-to-person interaction; therefore, it is difficult to capture or represent in explicit form». There is tacit knowledge in the understandings, skills and philosophies developed by societies with long histories of interaction with their natural surroundings. For rural and indigenous peoples, local knowledge informs decision-making about fundamental aspects of day-to-day life. This knowledge is integral to a cultural complex that also encompasses language, systems of classification, resource use practices, social interactions, ritual and spirituality (UNESCO 2014). Discussing tacit teaching, Miton and DeDeo (2022) point out that explicit knowledge can become tacit once the code required to express it is lost, and tacit knowledge can become explicit in cases where it can be codified. They found that tacit knowledge is acquired through socialization and sharing direct experience and interactions with others.

In a series of studies, Guarasci et al. (2008, 2010) explore the tacit knowledge of master craftsmen whose products are arguably the output of decades, sometimes centuries, of accumulated experiences, driven by what we consider brilliance or genius. They propose the creation of formalized workflows and

methods for the development of knowledge repositories necessary for guaranteeing continuity and succession related to specialized practices. In the field of goldsmith handcraftsmanship, for example, they identified a strong need to represent the regularly used knowledge and know-how of master craftsmen that allows for the production of exclusive and highly valuable products, without the need to describe and structure the whole production cycle.

According to Polanyi (1966, 17) true knowledge lies in our ability to use it. How can we use the brilliance of a master? Chess players, stresses Polanyi, enter into a master's spirit by rehearsing the games the master played to discover what the master had in mind, in an effort to acquire that tacit knowledge (Polanyi 1966, 30). We propose that a more direct path is needed to reliably access brilliance.

The multimodal analysis proposed in this paper allows for a deep exploration of the "why": why the chess master spends untold hours to study and improve, why the master goldsmith in Calabria keeps visualizing and creating works of beauty, what contributes to their brilliance, to their genius. The "why" is the most opaque part of tacit knowledge and its transmission is made possible by a thorough analysis of not only knowledge as linguistic content but also the manner in which it is expressed by the expert.

Language is the standard medium of knowledge documentation and transfer, but representing an individual's brilliance cannot rely on traditional means to study and preserve this tacit knowledge. Furthermore, tacit knowledge, when transferred, is often shared through face-to-face interactions. To capture the multidimensionality of genius and its expression requires a holistic approach to language analysis beyond a strictly unimodal view.

3. Language is multimodal

Traditional linguistic analysis is often unimodal, focusing on information expressed in a single modality and overlooking other linguistic and paralinguistic information that might be simultaneously present in other modalities. For instance, a mainstream semantic analysis might center around linguistic units such as morphemes, words, phrases, or constructions while idealizing away from non-syntactic information that might nevertheless contribute to the meaning of an utterance. In face-to-face interaction, though, speakers and hearers simultaneously use and integrate visuospatial and speech modalities during communication. If the analysis of language privileges a single channel, this ignores the wealth of additional information available when language is used face-to-face, including modulation of pitch and intonation, hand gestures, eye gaze, facial expression, and body posture. These are not just add-ons to the speech signal but are highly integrated into producing and understanding utterances. Based on this observation, there has been a growing movement to

analyze language as a multimodal phenomenon that is expressed simultaneously in visual and auditory channels (Holler and Levinson 2019; Perniss 2018).

Supporting this view that language is a multimodal phenomenon is abundant evidence that hearers use information from multiple modalities when understanding language. The classic example of this is the McGurk effect (McGurk and MacDonald 1976; MacDonald and McGurk 1978), where a voice saying [ba] is paired with a face articulating [ga], resulting in the perception of [da] in order to reconcile contradictory auditory and visual information. Visual information is used in other ways as well; for instance, speech is processed, interpreted, and even produced faster when facial articulatory movements can be seen (van Wassenhove, Grant, and Poeppel 2005; Drijvers and Holler 2023).

In addition to the visual modality corresponding to the face and mouth, there is evidence that co-speech gestures are also exploited by listeners and speakers. Adult speakers often use gesture when speaking (McNeill 1992; 2005; Kendon 2004; Goldin-Meadow 2003; Clark 1996), while children produce gestures even before speech, and the presence of gesture predicts certain linguistic milestones in language acquisition (see Goldin-Meadow 2014 for an overview). Gestures that mismatch a prior spoken word or are semantically incongruent with the spoken word show an N400 effect, an event-related potential (ERP) component in electroencephalogram (EEG) studies that is argued to index semantic anomaly (Kutas and Federmeier 2011). Listeners are sensitive to the temporal overlap between co-speech gestures and speech (McNeill 1992; Church, Kelly, and Holcomb 2013; Loehr 2007), and the integration of gesture with speech is partially determined by the emotional valence of the utterance (Levy and Kelly 2020; Kelly and Tran 2023).

Information from the visual modality can be linked to the utterance syntactically, semantically, and pragmatically. For instance, speakers are better at verbally communicating a two-dimensional shape to another speaker when allowed to use hand gestures (Graham and Argyle 1975), with the effect more pronounced when the shape was designed to be more difficult to communicate. This shows gesture may carry semantic information that can be understood by listeners. Iconic gestures contribute to the meaning of an utterance as well. A person might say "I didn't see it coming" while pantomiming the image of a car blind-siding them. Kelly et al. (1999) show that participants, when asked to summarize what was said, include in their summaries information that was contributed solely through iconic gestures. Co-speech gestures also contribute to how utterances cohere in a discourse (Lascarides and Stone 2009).

Some types of phonetic information can be integrated into an utterance in ways that are troublesome for mainstream approaches. For instance, formal semantics often identifies morphemes as the unit of semantic composition, but lengthening a vowel can sometimes signal semantic information that's not

strictly expressed by the morpheme itself (e.g., *it was a looong movie*), complicating traditional approaches to compositionality. Phonetic information is also used to convey pragmatic information such as irony and sarcasm, which are not necessarily found in the morphemic content of the utterance. Bryant and Fox Tree (2002) extracted sarcastic and non-sarcastic utterances from radio talk shows and presented them to experiment participants; participants were unable to distinguish which utterances were sarcastic or non-sarcastic based on the text of the utterance but could do so when the sound recordings were played. Filled pauses, like *uh* and *um*, control the flow of the discourse (Clark and Fox Tree 2002; Cossavella and Cevasco 2021), again operating distinctly from the morphemes in the utterance. And while the precise mapping between emotions and prosody is not perfectly understood (Larrouy-Maestri, Poeppel, and Pell 2024), experiments have shown that emotions and attitudes can be recognized based on the quality of voice alone (Pell 2006; Dara, Monetta and Pell 2008; Monetta, Cheang and Pell 2008). These examples show that we might gain better insight into language by setting aside assumptions about what can contribute to semantics and pragmatics, namely assumptions that understanding is based purely on words or morphemes alone.

This is a small sampling of the literature that shows that paralinguistic features are closely linked to speech, but it makes a case that language is expressed through multiple modalities simultaneously. Importantly, as Massaro (1987) points out, not only are these paralinguistic features used to encode part of the message, but they are also used in the interpretation of the message. There has therefore been a growing movement for corpora of spoken communication to include more than just transcripts of the spoken word, as transcripts cannot adequately capture the contribution of paralinguistic cues in verbal communication.

Examples of multimodal corpora that go beyond words are, among others, the SmartKom Corpus (Schiel, Steininger and Türk 2002), the Augmented Multi-party Interaction Corpus (Carletta et al. 2005), and the HuComTech Corpus (Abuczki and Ghazaleh 2013). These corpora include annotations at a variety of levels, including speech transcription, dialogue acts, group activity, topics and subtopics, abstractive and extractive summaries, where participants focused their attention, emotions expressed by facial expressions, emotion and physical cues, prosody, speech, intonation, head and hand gestures, hand shapes, movements, posture, and touches. Annotations such as these, in conjunction with annotations for pitch and body movement, have also been used to contribute to the development of multimodal sentiment analysis, which is used to automatically detect the positive, negative or neutral expression of emotion in a variety of modalities, as opposed to only texts (Gandhi et al. 2023).

As part of the movement for multimodal analyses of communication, we take the approach that if we want to understand the motivations behind brilliant people or what generates brilliance more generally, focusing solely on their words is not sufficient, as we may miss aspects of their language that signal precisely those motivations we are looking for. In this article we propose a method we believe can enable the discovery of what underpins brilliance. In the following section, we describe our approach to creating an annotated multimodal corpus that can be used to study the mindsets of exceptional individuals.

4. Methodology for discovering brilliance

4.1. In-depth structured interviews

In-depth structured interviews were conducted with 76 high-achieving individuals via video conference, obtaining verbal permission to record both video and audio. Participants were asked to prepare a list of accomplishments and were prompted to choose one accomplishment on which to focus during the interview. Participants were led through a question-answer protocol over a period of roughly 60-90 minutes.

Following the structured interviews, interviewers provided participants with a visual schematic of key excerpts of their own speech that represent the mindset components that lead to brilliance (see Yeramyan and Jackson 2014 for a description of the structured interview method and training). These excerpts are henceforth referred to as "genius excerpts." As described in white papers by Yeramyan and Jackson (2014; 2015), approximately ten thousand genius excerpts extracted from interview transcripts and notes were classified by a team of linguists based on a lexical semantic analysis of the text. However, the identified classes did not account for everything the participants expressed during their interviews. The themes displayed in Table 1 are derived from subsequent efforts to complement the text-based analysis (explicit knowledge) with a holistic analysis of the interview experience (tacit knowledge), which resulted in a more refined thematic grouping. These seven major themes characterize the mindsets that underlie the brilliant accomplishments of high-achieving individuals.

possibility / creation / choice / freedom
purpose / contribution / world impact / legacy
extremeness / intensity / obsession / above and beyond
curiosity / discovery / learning / growing
at-stakeness[1] / conviction / resolve / perseverance
joy / excitement / exhilaration / spirit
togetherness / relationship / connection / love

Table 1. Seven themes of brilliance (Gap International 2021).

As Braun, Goudbeek, and Krahmer (2020) point out, there are disparities between self- and other-annotation. They conclude that discrepancies between authors' intentions and readers' perceptions might indicate a mismatch between the authors' intended affect and the affect experience of others. To minimize this discrepancy, participants were prompted to classify their own genius excerpts according to these seven themes; they were able to select up to three of the seven themes for each genius excerpt.

4.2. Interview Annotation

4.2.1. *Full interview annotation*

The first phase of analysis focused on the theme of extremeness / intensity / obsession / above and beyond. Four of the structured interviews were selected for an in-depth examination. One annotator watched the interviews in their entirety and selected sections where they noted the theme. A team of annotators then watched these sections and collectively identified the linguistic and paralinguistic features indicating the presence of the theme. The team aligned on which features were most salient and categorized them into linguistic categories: lexical, phonetic, rhetorical, syntactic; and paralinguistic categories: eye gaze, facial expressions, body movement, and hand gestures. The EUDICO Linguistic Annotator (ELAN) video and audio annotation software (The Language Archive 2023) was selected for annotation, and a tier was created for each category. Throughout this annotation phase, there were between four and six annotators (as some annotators joined the project after the initial phase).

Once the feature tags were finalized, the annotators individually watched each interview in full and tagged the features in sections they identified as relevant to the theme of extremeness / intensity / obsession / above and beyond. All tags were applied at the word or phrase level. In addition to an-

[1] Here "at-stakeness" is a neologism used to mean the condition of having something at stake.

notating features of this theme, annotators also used a dedicated ELAN tier to flag sections where they perceived any of the other six themes of brilliance to be present in the interviewee's speech, with the intent to use this information in the following phase of research. All practical instructions on implementing feature tags were documented for reference in a set of annotation guidelines.

For two lexical features of the theme, absolute and intensifier, automatic tagging was used to aid annotators as they navigated through the entire interview. The team collectively composed a word list with relevant lexemes. For example, the list of absolutes included lexemes like *any, all, every, each, always, never* and the list of intensifiers included lexemes such as *very, really, highly, so, super*. The algorithm tagged any instance of the lexemes, which produced considerable noise. For this reason, the automatically identified tags were not incorporated directly into the data set, but instead used as a tool to direct the annotator's attention to relevant tokens. The automatic tagging flagged all instances of the lexemes on a dedicated tier in ELAN, and the annotator tagged only relevant tokens on the lexical tier. For instance, an annotator would ignore an instance of *so* as a conjunction (e.g. *so I left*) but tag an instance of *so* as an intensifier (e.g. *so determined*).

In the first phase of annotation, the paralinguistic features were tagged both manually and automatically. A video analysis software tool based on open-source libraries was created to automatically detect and tag visible facial expressions, emotions, eye gaze, and body movements (e.g. *furrowed brow*, *surprise*, *anger*, *happiness*, *widen eyes*, *lean in*, *lean out*, etc.). As visual analysis software uses frame-by-frame analysis (24 frames per second), it was decided to record only the dominant tag during the pronunciation of a word segment (e.g., if the enunciation of a word took 0.5 seconds, among the 12 tags generated for the 12 frames, only the most frequent tag would be reported for that time interval). Still, there was a significant overproduction of tags which caused more noise than signal for the purpose of this research. A persistent issue with automatic annotation was a high error rate (e.g., the software would interpret the V-shape of certain glasses as a furrowed brow and therefore tag it as *angry*). For this reason, although automatic annotations of visual cues were produced for all videos, they were not included in the final analysis.

4.2.2. *Full interview annotation review*

After the initial annotation, each structured interview was reviewed by another annotator. The reviewer watched the interview in full, checking all annotations as well as unannotated sections. The reviewer made notes in a separate review tier about any proposed changes and flagged newly added annotations.

The reviewer and annotator then met to discuss disagreements or new annotations made by the reviewer and align on the final annotations. Any cases for which the reviewer and annotator could not reach an agreement were brought to the whole annotation team for a final decision. In addition to the changes agreed upon by the reviewer and annotator, there was a list of sanctioned changes and additions which the reviewer could make without consulting the annotator. Sanctioned changes and additions included those made to reflect updates to annotation guidelines or new annotations on the phonetic tier.

4.2.3. *Partial interview annotation*

A second approach was taken for analysis of the remaining six themes, using portions of each of the 76 structured interviews rather than analyzing the entire interviews. A research cycle was designed to study one theme of brilliance at a time, as follows.

The team began with preliminary analysis to determine the linguistic features of the theme at hand using the genius excerpts, as classified by the participants. For this analysis, annotators used a sample of ten genius excerpts from each participant's interview to identify the linguistic features for the given theme. The team aligned on which features were most salient and categorized them into linguistic categories: lexical, phonetic, rhetorical, semantic and syntactic. The paralinguistic categories from the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme were carried over to this analysis: eye gaze, facial expressions, body movement, and hand gestures. For each theme of brilliance, all feature tags were defined and documented in annotation guidelines for reference during annotation.

Once the feature tag set was defined, a preliminary annotation phase was carried out with a subset of the structured interviews to evaluate the relevance of the tags and to note considerations in implementation. Each annotator completed annotation of one randomly chosen interview, annotating 3-4 sections using each of three annotation passes (as described below). In an alignment meeting, annotators proposed changes to the feature tags, which were agreed upon by the whole team and documented in the annotation guidelines. After the feature tag sets were finalized, the team annotated all participant interviews using ELAN, meeting daily to resolve issues or disagreements on their use. Between six and ten annotators completed annotations for a given theme of brilliance.

Portions of each interview were selected by conducting three annotation passes for each interview. The first pass used each participant's genius excerpts for each theme, according to their thematic classification. The genius excerpts were located in the interviews using a matching algorithm; the team annota-

ted features in these sections and in the utterances immediately preceding and following. The second pass consisted of iterating through sections that had been previously flagged as corresponding to one of the six themes of brilliance, identified during the annotation of EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND described above.

The third pass focused on lexemes that were tagged automatically using a matching algorithm. The annotation team created lexical sets composed of lexemes that are semantically similar to the themes of brilliance. For example, the POSSIBILITY / CREATION / FREEDOM / CHOICE theme's lexical set included words like *opportunity, innovation, potential, invent.* These lexical sets were selected by the annotators based on review of the genius excerpts and word embedding techniques based on language generators. The first iteration of this matching algorithm used lemmas to detect multiple forms of a lexeme (e.g., *believe, believed, believing*). It was noted that the algorithm was missing target lexemes that were subsequently found by the annotator. A stemming approach eventually replaced the lemmatization approach to capture more lexemes, which yielded a more effective set of automatically tagged lexemes. Annotators then checked tagged lexemes and the surrounding utterances for the presence of the brilliance theme at hand.

Some features were identified across themes, and therefore their corresponding tags were carried over from annotation of one theme to the next. Two such features were the use of absolutes and intensifiers. The automatic tagging of absolutes and intensifiers, as described above for analysis of the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme, was also available as an ELAN tier to annotators to reference in all three annotation passes.

The three passes were effective in directing annotators' attention to sections of the interview relevant to the theme being studied. Taken together, the passes ensured that the selected portions were indeed relevant. Each annotation pass did at times direct annotators to sections that were already viewed in another pass; however, each pass also identified additional sections to annotate.

4.2.4. *Partial interview annotation review*

Each annotated interview was reviewed by another member of the annotation team. The reviewer checked all annotations and labeled their agreement for each annotation on a separate agreement tier using binary documentation: 1 for agreement and 0 for disagreement. The reviewer could also input additional annotations to the annotated sections.

The reviewer and annotator then met to discuss disagreements or new annotations made by the reviewer and align on the final annotations. Any cases for which the reviewer and annotator could not reach an agreement were brought to the whole annotation team for a final decision. In addition to the

changes agreed upon by the reviewer and annotator, there was a list of sanctioned changes and additions which the reviewer could make without consulting the annotator. Sanctioned changes included: those made to reflect updates to annotation guidelines or new annotations of tags that were well-defined and less subject to annotator judgement (e.g. phonetic features). In rare cases, newly defined feature tags that were straightforward in their use (e.g. questions) were also included as a sanctioned change. Any changes made, both new annotations and revisions to existing annotations, were counted and marked on a separate change tier.

Figure 1 lays out the major steps of the annotation process as conducted for the partial interview approach.

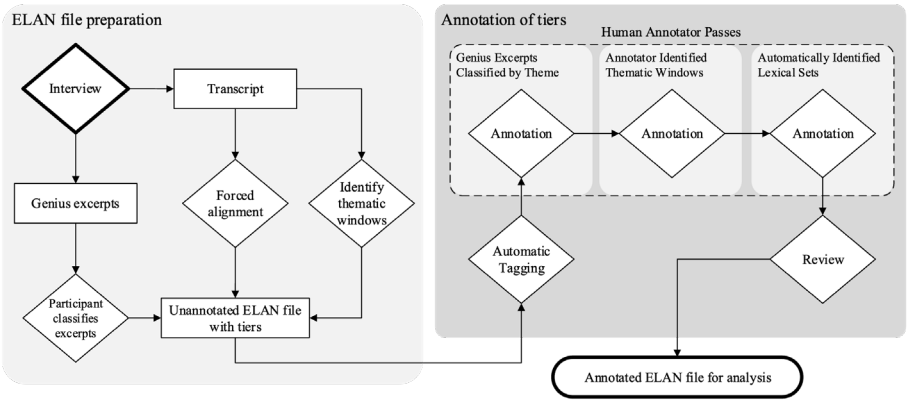


Figure 1: Visualization of the file preparation and annotation process, starting with the structured interview and ending with an annotated ELAN file.

4.2.5. *Partial interview annotation inter-rater reliability*

To measure the consistency and reliability of the subjective annotation system, the team used an inter-rater reliability metric that tests whether the annotator and reviewer's judgments were the same at a rate significantly higher than one would expect due to chance. It is commonly used in place of percent agreement as a way of evaluating the agreement rate while considering the occurrence of chance agreements.

The metric, Cohen's κ (kappa), uses the relative observed agreement among raters, and the hypothetical probability of chance agreement to evaluate the agreement rate in relation to the possibility of chance agreements (Cohen 1960). Table 2 shows the κ scores and the agreement rate between the annotators and the reviewers for each theme.

Brilliance theme	Cohen's κ	Agreement
possibility / creation / choice / freedom	0.478	95.8%
purpose / contribution / world impact / legacy	0.456	91.9%
curiosity / discovery / learning / growing	0.488	97.8%
at-stakeness / conviction / resolve / perseverance	0.439	89.2%
joy / excitement / exhilaration / spirit	0.463	93.2%
togetherness / relationship / connection / love	0.450	91%

Table 2: Kappa statistic and agreement rate by brilliance theme.

According to the guidelines set forth by Landis and Koch (1977), all of the κ statistics are within the Moderate Agreement range (0.4 to 0.6). This indicates that while there were disagreements about the subjective annotations, the reviewer and the annotator had identical judgments at a rate significantly higher than one would expect from chance. This suggests a relatively high degree of alignment among annotators regarding feature tags and their use.

4.2.6. *Interview annotation summary*

In summary, two distinct annotation processes were used; one in which the entire interview was viewed, and one in which only selected portions of the interview were viewed. In the first approach, all sections of the interview identified as relevant to the theme were annotated according to the annotator's discretion. In the second approach, the annotators used a more systematic approach to identify sections relevant to the theme, incorporating participant input to do so. The second approach allowed the annotation team to annotate the remaining themes in one sixth of the time that was required for the first approach, while the first provided annotators with a comprehensive view of the interview data. Future investigations may adopt the full interview annotation method for other themes in order to draw a direct comparison between annotation approaches and more fully evaluate the merits of each.

4.2.5 *Acoustic dimension of annotations*

Speech offers a natural and complex way to convey information about our emotional and mental state. In conversational or spontaneous speech, such information is often expressed unintentionally, yet small acoustic changes can create large perceptual effects heard by the listener. Many studies have been conducted to determine which acoustic correlates are significant in the study of emotional speech (Weninger et al. 2013; Kamiloğlu, Fischer, and Sauter 2020; Kacur et al. 2021). Similarly, acoustic correlates have been identified for

diagnosing various health conditions (Roark et al. 2011; Meilán et al. 2014; Rektorova et al. 2016; Sara et al. 2020). While there have been attempts to examine the acoustic correlates of complex expressions like charisma (Hirschberg and Rosenberg 2005), leadership (Weninger et al. 2012), and cooperation (Tognetti et al. 2020), little has been done to study the acoustic correlates of brilliance. In the present multimodal study, we aim to identify acoustic correlates of each brilliance theme by extracting acoustic features used in the previously mentioned studies. We hypothesize that acoustic patterns map onto clusters which can classify a particular brilliance theme.

Transcripts for each interview were obtained using the default Automatic Speech Recognition output in Video Text Tracks (VTT) (World Wide Web Consortium 2019) format, and later were improved using WhisperX (Bain et al., 2023) and underwent automatic text-to-speech alignment via Montreal Forced Aligner (MFA) (McAuliffe et al. 2017) using an American English pronunciation dictionary. Automatic alignment offers sufficiently accurate parsing of speech and pause regions for use in speech-based measures (Yuan et al. 2023). Output alignments resulted in 52 TextGrids with word-level and phone-level alignments for 76 speakers[2]. TextGrids with word-level alignments were imported into ELAN to reference during manual annotation. Speech disfluencies in conversational speech (i.e. hesitations, interruptions, filled pauses) and low recording quality created some misalignment between text and audio. However, these were manually corrected during annotation. Similarly, mistranscribed words were manually changed to the correct word in accordance with the audio; these corrected transcripts were imported from the annotation of one theme to the next. A separate 'Pause Tier' was manually generated in ELAN from empty intervals on the word-tier to ensure perceived pauses were annotated with time (ms) accuracy. Pause regions were defined as intervals where no acoustic signal was detected by MFA.

Our proposed method of annotation is based on existing research on emotions and voice, in which both acoustic features and listeners' perception is measured (Kamiloğlu, Fischer, and Sauter 2020; Dai, Fell, and MacAuslan 2009; Erickson et al. 2016; Kacur et al. 2021). Prior to extracting acoustic features, the data was manually annotated in ELAN for perceptual effects of acoustics in a "Phonetic Tier". These perceptual tags[3] were simply defined as: *pitch change*, *emphatic stress*, *rhythm*, and *pause*. Perceptual annotation of the data offered an indication if the brilliance theme in question was acoustically perceived by the annotator or not. Thus, tag labels were not meant to have a

---

[2] The difference in number of TextGrid files and number of speakers is due to multiple speakers per interview. Each TextGrid file corresponds to one interview, some containing multiple speakers.

[3] Perceptual tags for acoustics were the same for all themes except for EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND, as this was the first theme annotated. Therefore tags for EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND were simplified to the current tags.

one-to-one relationship with the acoustic feature extracted for that region, but rather an indication that the given brilliance theme was phonetically marked compared to the surrounding speech.

Our approach to acoustic feature extraction includes features reported in the literature as relevant in emotional speech detection and speech used to detect health conditions. Acoustic features across four domains will be extracted in Praat (Boersma and Weenink 2024), namely (i) pitch, (ii) voice quality, (iii) intensity, and (iv) rate (temporal measures) (Tab. 3). Multiple acoustic measures across these domains have proven best for classification results (Kacur et al. 2021; Kamiloğlu, Fischer, and Sauter 2020). The table in Appendix III shows all acoustic features to be extracted (*fundamental frequency*, *formant frequencies*, *harmonics to noise ratio*, *intensity and amplitude*, *shimmer*, *speech rate*, *articulation rate*, and *pauses*) and their corresponding tags, perceptual correlates, and definitions. Praat scripts for extracting acoustic features were adapted for the data to iterate on a by-speaker basis and parameters were adjusted to account for variables shown affect acoustics (i.e. gender). These scripts were run on small batches of the Genius Excerpts to see preliminary acoustic trends and adjust script parameters. These scripts iterate over voiced regions of speech unless the acoustic feature specifically calls for unvoiced regions.

5. From qualitative to quantitative annotation analysis: a preliminary sample

After the annotation process, certain linguistic and paralinguistic features were observed to appear more prominently in some themes over others. The annotation process itself offered ongoing qualitative insights into the motivations and emotional triggers that generate brilliance, such as what features should we look for and how should they be tagged. That very process created the basis for a hypothesis. Based on those hypotheses, we proceeded with a quantitative analysis of the generated annotations. Below is a preliminary analysis of two quantitative features, one from the lexical domain and one from the acoustic domain.

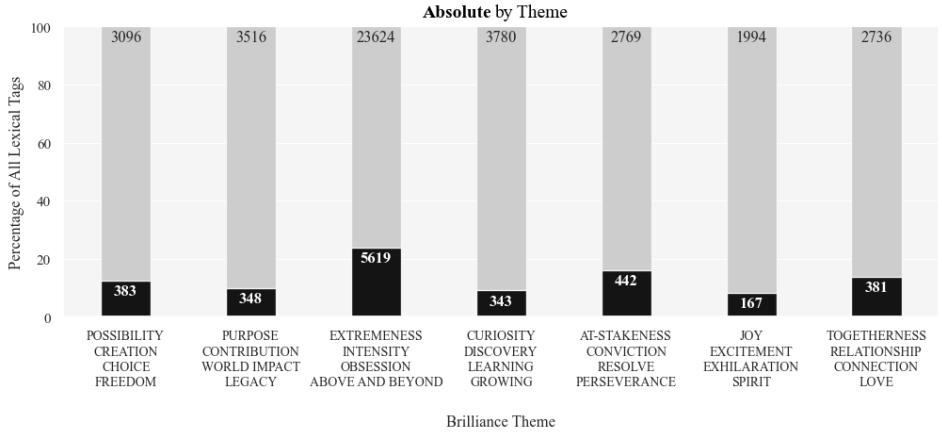


Figure 2: Absolute lexical tag proportion by theme.

Brilliance theme	Proportion of *absolute* tags to Lexical tags	Total *absolute* tags
POSSIBILITY / CREATION / CHOICE / FREEDOM	0.124	383
PURPOSE / CONTRIBUTION / WORLD IMPACT / LEGACY	0.099	348
EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND	0.238	5619
CURIOSITY / DISCOVERY / LEARNING / GROWING	0.091	343
AT-STAKENESS / CONVICTION / RESOLVE / PERSEVERANCE	0.160	442
JOY / EXCITEMENT /EXHILARATION / SPIRIT	0.084	167
TOGETHERNESS / RELATIONSHIP / CONNECTION / LOVE	0.129	381
Weighted mean proportion for all themes	0.206	7683

Table 3: *absolute* lexical tags as a proportion of all lexical tags, organized by brilliance theme.

Table 3 shows the number of *absolute* as a proportion of the total lexical tags for each theme. The *absolute* tag was used to indicate the presence of lexemes that indicated an all or nothing meaning (e.g. *always, every time, never, total*). The hypothesis that absolutes were a more salient feature of the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme than the other themes was tested using a one-sample proportion test.

The test measured whether there was a statistically significant difference between the proportion of *absolute* to all lexical tags in the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme (0.238) than for the

overall weighted mean proportion of *absolute* to all lexical tags (0.206). The one-sample proportion test revealed that the higher-than-average proportion of *absolute* for the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme was significant $\chi^2 = 141.1$, $p < 0.0001$. This indicates that in sections where the EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND theme was noted by annotators, the interviewees were more likely to use absolutes in their speech, suggesting that the use of absolutes is an indicative feature of the expression of EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND.

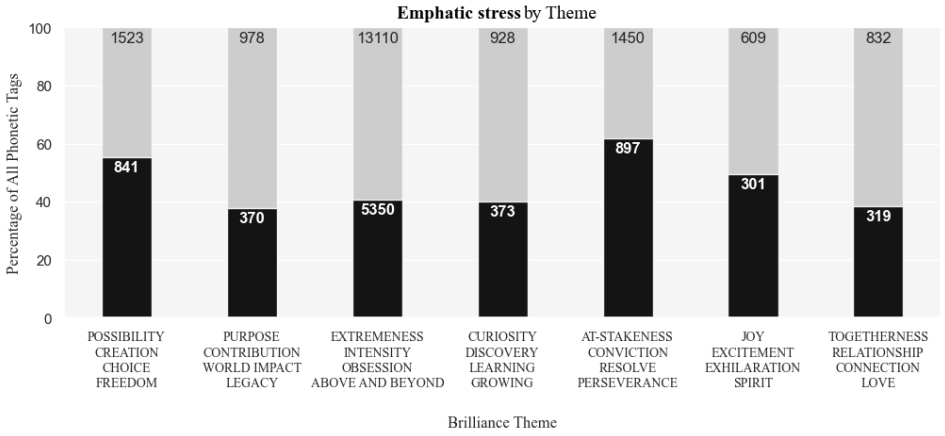


Figure 3: *emphatic* stress tag proportion by theme.

Brilliance theme	Proportion of *emphatic stress* tags to Phonetic tags	Total *emphatic stress* tags
POSSIBILITY / CREATION / CHOICE / FREEDOM	0.552	841
PURPOSE / CONTRIBUTION / WORLD IMPACT / LEGACY	0.378	370
EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND	0.408	5350
CURIOSITY / DISCOVERY / LEARNING / GROWING	0.402	373
AT-STAKENESS / CONVICTION / RESOLVE / PERSEVERANCE	0.619	897
JOY / EXCITEMENT /EXHILARATION / SPIRIT	0.494	301
TOGETHERNESS / RELATIONSHIP / CONNECTION / LOVE	0.383	319
Weighted mean proportion for all themes	0.445	8451

Table 4: *emphatic stress* tags as a proportion of all Phonetic tags, organized by brilliance theme.

The *emphatic stress* tag was used when the annotator perceived word-level stress used to express the theme, beyond the normal lexical stress. Table 4

shows the *emphatic stress* tags as a proportion of the total Phonetic tags for each theme.

The *emphatic stress* tag was prevalent throughout the seven themes. Analysis showed it to be most prominent in sections noted for the AT-STAKENESS / CONVICTION / RESOLVE / PERSEVERANCE theme. A one-sample proportion test was performed to determine if there was a statistically significant difference between the proportion of *emphatic stress* in the AT-STAKENESS / CONVICTION / RESOLVE / PERSEVERANCE theme (0.619) and the overall weighted mean proportion of *emphatic stress* tags (0.445). The one-sample proportion test indicated that there is a statistically significant difference, $\chi^2 = 176.27$, $p < 0.0001$. The results of this test show that the difference from the average proportion of *emphatic stress* across all themes is statistically significant. This suggests that *emphatic stress* was more likely to appear in sections where the AT-STAKENESS / CONVICTION / RESOLVE / PERSEVERANCE theme was noted, as compared to the other themes.

6. Conclusions and future work

The examination of what causes human brilliance is an area of inquiry which needs methodologies that are not only useful for explicit forms of knowledge, but can also adequately capture tacit knowledge. The implementation of a multimodal annotation approach for the analysis of structured interviews with individuals sharing their brilliance has revealed the complex interaction of cognitive factors that underpin exceptional accomplishments. The diversity of these factors requires the expansion of traditional analytical frameworks to embrace the different data types generated in conversation, such as textual transcripts and acoustic and visual data. In defining the methodology proposed here, every effort was made to minimize annotators' perception biases and to maximize the participants' intention through linguistic and behavioral analysis of the interviewee.

The data generated by this annotation process enables a systematic uncovering of patterns related to mindset and attitudes, while maximizing the codification of tacit knowledge within narrative accounts of brilliance. By employing quantitative measures of the cooccurrence of themes of brilliance with factors such as acoustic features, contextually relevant gestures, and semantic clusters, a more structured examination of the attributes associated with brilliance is made possible.

Future explorations into the domain of brilliance may benefit significantly from the incorporation of artificial intelligence (AI) and the automation of analytical processes. AI's capabilities in managing extensive datasets and identifying complex patterns could greatly augment the depth and efficiency of analysis. Automating elements of the annotation process, including emotional

tone assessment, thematic tagging through natural language processing, and pattern identification, could streamline research methodologies, allowing for a focus on more interpretive aspects of analysis.

The increased attention to the role of multimodal interaction with Large Language Models enables the development of annotation tools specifically designed for multimodal data, aiming to improve both the precision and user experience for researchers. We are also considering the integration of machine learning algorithms to refine pattern detection capabilities and to enhance the predictability of traits associated with brilliance based on annotated interviews. Automation combined with careful human fine-tuning will enable longitudinal research studies to capture the developmental trajectory of brilliance, providing insight into its dynamism and adaptability. Future research directions also include expanding and balancing the participant pool across various disciplines, cultural backgrounds, and personal histories to ensure a more inclusive representation of brilliance.

In summary, the scholarly pursuit of understanding human brilliance necessitates an ongoing commitment to methodological innovation and interdisciplinary exploration. The multimodal annotation method proposed within this study offers a foundational approach for dissecting the intricate web of brilliance. With the prospective integration of AI and automation, research into human potential is poised for advancements that promise richer insights and more nuanced understandings of how brilliance manifests and could be cultivated across human activity.

References

Abuczki, Ágnes, and Esfandiari Baiat Ghazaleh. 2013. "An overview of multimodal corpora, annotation tools and schemes." *Argumentum* 9(1): 86-98. https://core.ac.uk/download/pdf/161005866.pdf.

Bain, Max, Jaesung Huh, Tengda Han, and Andrew Zisserman. 2023. "WhisperX: Time-accurate speech transcription of long-form audio." In *Proceedings of Interspeech* 2023*,* 4489-93. https://doi.org/10.21437/Interspeech.2023-78.

Boersma, Paul, and Weenink, David. 2024. *Praat: doing phonetics by computer* [Computer program]. V6.1.09. http://www.praat.org/.

Braun, Nadine, Martijn Goudbeek, and Emiel Krahmer. 2020. "Emotional Words-The Relationship of Self-and Other-Annotation of Affect in Written Text." *Cognitive Science:* 430-36.

Bryant, Gregory A., and Jean E. Fox Tree. 2002. "Recognizing verbal irony in spontaneous speech." *Metaphor and Symbo*l 17(2): 99-117. https://doi.org/10.1207/S15327868MS1702_2.

Carletta, Jean, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, et al. 2005. "The AMI meeting corpus: A pre-announcement." In *International workshop on machine learning for multimodal interaction*: 28-39. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11677482_3.

Casonato, R., and K. Harris. 1999. "Can an Enterprise Really Capture 'Tacit Knowledge': We answer two top questions on knowledge management from the Electronic Workplace 1999 Conference." *Gartner Group Research Note Select Q&A* 16.

Church, R. Breckinridge, Spencer Kelly, and David Holcombe. 2013. "Temporal synchrony between speech, action and gesture during language production." *Language, Cognition and Neuroscience* 29(3): 345-54. https://doi.org/10.1080/01690965.2013.857783.

Clark, Herbert H. 1996. *Using language*. Cambridge University Press. https://doi.org/10.2277/0521561582.

Clark, Herbert H., and Jean E. Fox Tree. 2002. "Using *uh* and *um* in spontaneous speaking." *Cognition* 84(1): 73-111. https://doi.org/10.1016/S0010-0277(02)00017-3.

Cohen, Jacob. 1960. "A coefficient of agreement for nominal scales." *Educational and psychological measurement* 20(1): 37-46. http://dx.doi.org/10.1177/001316446002000104.

Cossavella, Francisco, and Jazmín Cevasco. 2021. "The importance of studying the role of filled pauses in the construction of a coherent representation of spontaneous spoken discourse." *Journal of Cognitive Psychology* 33(2): 172-86. https://doi.org/10.1080/20445911.2021.1893325.

Dai, Keshi, Harriet Fell, and Joel MacAuslan. 2009. "Comparing Emotions Using Acoustics and Human Perceptual Dimensions." In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* CHI EA '09: 3341-46. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1520340.1520483.

Dampney, Kit, Peter Busch, and Debbie Richards. 2002. "The meaning of tacit knowledge." *Australasian Journal of Information Systems* 10(1): 3-13. https://doi.org/10.3127/ajis.v10i1.438.

Dara, Chinar, Laura Monetta, and Marc D. Pell. 2008. "Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions." *Brain Research* 1188: 100-11. https://doi.org/10.1016/j.brainres.2007.10.034.

Drijvers, Linda, and Judith Holler. 2023. "The multimodal facilitation effect in human communication." *Psychonomic Bulletin & Review* 30*:* 792-801. https://doi.org/10.3758/s13423-022-02178-x.

Erickson, Donna, Chunyue Zhu, Shigeto Kawahara, and Atsuo Suemitsu. 2016. "Articulation, Acoustics and Perception of Mandarin Chinese Emotional Speech." *Open Linguistics* 2(1): 620-35. https://doi.org/10.1515/opli-2016-0034.

Gandhi, Ankita, Kinjal Adhvaryu, Soujanya Poria, Erik Cambria, and Amir Hussain. 2023. "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions." *Information Fusion* 91: 424-44. https://doi.org/10.1016/j.inffus.2022.09.025.

Gap International. 2021. "The Leveraging Genius Conference 2021." Online, February 22-24, 2021, https://m.eventsinamerica.com/events/the-leveraging-genius-conference-2021/retail-wholesale/ecommerce/5obagztduc3696mz.

Goldin-Meadow, Susan. 2003. *Hearing gesture: How our hands help us think.* Harvard University Press.

Goldin-Meadow, Susan. 2014. "Widening the lens: What the manual modality reveals about language, learning and cognition." *Philosophical Transaction of the Royal Society B* 369: 20130295. https://doi.org/10.1098/rstb.2013.0295.

Graham, Jean Ann, and Michael Argyle. 1975. "A cross-cultural study of the communication of extra-verbal meaning by gestures." *International Journal of Psychology* 10(1): 57-67. https://doi.org/10.1080/00207597508247319.

Guarasci, Roberto, Antonietta Folino, Alessia Cosentino, Elena Cardillo, and Maria Taverniti. 2008. "Gestion et formalisation de la connaissance tacite. " In *Sistemes d'Information & Intelligence economique* 1: 564-76. IHE éditions.

Guarasci, Roberto, Elena Cardillo, Antonietta Folino, and Maria Taverniti. 2010. "Multilingual taxonomic and terminological structures of a domain." In *Globalization and the Management of Information Resources - Papers from the International Conference. Sofia, Bulgaria, 12-14 November 2008*, 508-18. St. Kliment Ohridski University Press.

Hirschberg, Julia Bell, and Andrew Rosenberg. 2005. "Acoustic/Prosodic and Lexical Correlates of Charismatic Speech." In *Proceedings of Eurospeech 2005*. https://doi.org/10.7916/D8Q52Z3Z.

Holler, Judith, and Stephen C. Levinson. 2019. "Multimodal language processing in human communication." *Trends in Cognitive Sciences* 23(8): 639-52. https://doi.org/10.1016/j.tics.2019.05.006.

Kacur, Juraj, Boris Puterka, Jarmila Pavlovicova, and Milos Oravec. 2021. "On the Speech Properties and Feature Extraction Methods in Speech Emotion Recognition." *Sensors* 21(5): 1888. https://doi.org/10.3390/s21051888.

Kamiloğlu, Roza G., Agneta H. Fischer, and Disa A. Sauter. 2020. "Good Vibrations: A Review of Vocal Expressions of Positive Emotions." *Psychonomic Bulletin & Review* 27(2): 237–65. https://doi.org/10.3758/s13423-019-01701-x.

Kelly, Spencer D., and Quang-Anh Ngo Tran. 2023. "Exploring the emotional functions of co-speech hand gesture in language and communication." *Topics in Cognitive Science* 00: 1-23. https://doi.org/10.1111/tops.12657.

Kelly, Spencer D., Dale J. Barr, Ruth Breckinridge Church, and Katheryn Lynch. 1999. "Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory." *Journal of Memory and Language* 40: 577-92. https://doi.org/10.1006/jmla.1999.2634.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

Kutas, Marta, and Kara D. Federmeier. 2011. "Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP)." *Annual Review of Psychology* 62: 621-47. https://doi.org/10.1146/annurev.psych.093008.131123.

Landis, J. Richard, and Gary G. Koch. 1977. "The measurement of observer agreement for categorical data." *Biometrics*: 159-74. http://dx.doi.org/10.2307/2529310.

Larrouy-Maestri, Pauline, David Poeppel, and Marc D. Pell. 2024. "The sound of emotional prosody: Nearly 3 decades of research and future directions." *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916231217722.

Lascarides, Alex, and Matthew Stone. 2009. "Discourse coherence and gesture interpretation." *Gesture* 9(2): 147-80. https://doi.org/10.1075/gest.9.2.01las.

Levy, Rachel S., and Spencer D. Kelly. 2020. "Emotion matters: The effect of hand gesture on emotionally valenced sentences." *Gesture* 19(1): 41-71. https://doi.org/10.1075/gest.19029.lev.

Loehr, Daniel. 2007. "Aspects of rhythm in gesture and speech." *Gesture* 7(2): 179-214. https://doi.org/10.1075/gest.7.2.04loe.

MacDonald, John, and Harry McGurk. 1978. "Visual influences on speech perception processes." *Perception and Psychophysics* 24(3): 253-57. https://doi.org/10.3758/BF03206096.

Massaro, Dominic W. 1987. *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, New Jersey: Lawrence Erlbaum.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. "Montreal Forced Aligner: trainable text-speech alignment using Kaldi." In *Proceedings of the 18th Conference of the International Speech Communication Association:* 498-502. https://doi.org/10.21437/Interspeech.2017-1386.

McGurk, Harry, and John MacDonald. 1976. "Hearing lips and seeing voices." *Nature* 264: 746-48. https://doi.org/10.1038/264746a0.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

McNeill, David. 2005. *Gesture and thought.* University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001.

Meilán, Juan José G., Francisco Martínez-Sánchez, Juan Carro, Dolores E. López, Lymarie Millian-Morell, and José M. Arana. 2014. "Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia?" *Dementia and Geriatric Cognitive Disorders; Basel* 37(5-6): 327-34. https://doi.org/10.1159/000356726.

Monetta, Laura, Henry S. Cheang, and Marc D. Pell. 2008. "Understanding speaker attitudes from prosody by adults with Parkinson's disease." *Journal of Neuropsychology* 2: 415-30. https://doi.org/10.1348/174866407X216675.

Miton, Helena, and Simon DeDeo. 2022. "The cultural transmission of tacit knowledge." *Journal of the Royal Society Interface* 19(195): 20220238. https://doi.org/10.1098/rsif.2022.0238.

Olomolaiye, Anthony, and Charles Egbu. 2005. "Tacit vs. explicit knowledge - the current approaches to knowledge management." In *Second Scottish Conference for Postgraduate Researchers of the Built and Natural Environment (Probe)*: 503–11. https://www.irbnet.de/daten/iconda/CIB10682.pdf.

Pell, Marc D. 2008. "Cerebral mechanisms for understanding emotional prosody in speech." *Brain and Language* 96(2): 221-34. https://doi.org/10.1016/j.bandl.2005.04.007.

Perniss, Pamela. 2018. "Why we should study multimodal language." *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.01109.

Polanyi, Michael. 1958. *Personal Knowledge: Towards a Post - Critical Philosophy*. London: Routledge and Kegan Paul.

Polanyi, Michael. 1966. *The Tacit Dimension.* Garden City, NY: Doubleday & Company Inc.

Rektorova, Irena, Jiri Mekyska, Eva Janousova, Milena Kostalova, Ilona Eliasova, Martina Mrackova, Dagmar Berankova, Tereza Necasova, Zdenek Smekal, and Radek Marecek. 2016. "Speech Prosody Impairment Predicts Cognitive Decline in Parkinson's Disease." *Parkinsonism & Related Disorders* no. 29 (August): 90-95. https://doi.org/10.1016/j.parkreldis.2016.05.018.

Roark, Brian, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffrey Kaye. 2011. "Spoken Language Derived Measures for Detecting Mild Cognitive Impairment." In *IEEE Transactions on Audio, Speech, and Language Processing* 19(7): 2081–90. https://doi.org/10.1109/TASL.2011.2112351.

Sara, Jaskanwal Deep Singh, Elad Maor, Barry Borlaug, Bradley R. Lewis, Diana Orbelo, Lliach O. Lerman, and Amir Lerman. 2020. "Non-Invasive Vocal Biomarker Is Associated with Pulmonary Hypertension." *PLOS ONE* 15(4): e0231441. https://doi.org/10.1371/journal.pone.0231441.

Schiel, Florian, Silke Steininger, and Ulrich Türk. 2002. "The SmartKom Multimodal Corpus at BAS." In *Proceedings of the International Language Resources and Evaluation Conference (LREC):* 200-06. http://www.lrec-conf.org/proceedings/lrec2002/pdf/49.pdf.

Simonton, Dean Keith. 2016. "Reverse engineering genius: historiometric studies of superlative talent." *Annals of the New York Academy of Sciences* 1377 (1): 3-9. https://doi.org/10.1111/nyas.13054.

The Language Archive. 2023. *ELAN* [Computer program] V.6.7. Max Planck Institute for Psycholinguistics. https://archive.mpi.nl/tla/elan.

Tognetti, Arnaud, Valerie Durand, Melissa Barkat-Defradas, and Astrid Hopfensitz. 2020. "Does He Sound Cooperative? Acoustic Correlates of Cooperativeness." *British Journal of Psychology* 111(4): 823–39. https://doi.org/10.1111/bjop.12437.

UNESCO. 2014. "Inventory-making: a cumulative in-depth study of periodic reports." UNESCO Intangible Cultural Heritage. Accessed April 8, 2024. https://ich.unesco.org/en/focus-on-inventory-making-2014-00876.

van Wassenhove, Virginie, Ken W. Grant, and David Poeppel. 2005. "Visual speech speeds up the neural processing of auditory speech." In *Proceedings of the National Academy of Sciences* 102 (4): 1181-86. https://doi.org/10.1073/pnas.0408949102.

Weninger, Felix, Florian Eyben, Björn Schuller, Marcello Mortillaro, and Klaus Scherer. 2013. "On the Acoustics of Emotion in Audio: What Speech, Music, and Sound Have in Common." *Frontiers in Psychology* 4: 292. https://doi.org/10.3389/fpsyg.2013.00292.

Weninger, Felix, Jarek Krajewski, Anton Batliner, and Björn Schuller. 2012. "The Voice of Leadership: Models and Performances of Automatic Analysis in Online Speeches." *IEEE Transactions on Affective Computing* no. 3 (October): 496-508. https://doi.org/10.1109/T-AFFC.2012.15.

World Wide Web Consortium (W3C). 2019. "WebVTT: The Web Video Text Tracks Format." https://www.w3.org/TR/webvtt1/.

Yeramyan, Pontish, and Eric Jackson. 2014. "Tapping into Genius." https://www.gapinternational.com/insights/details/index/tapping-into-genius.

Yeramyan, Pontish, and Eric Jackson. 2015. "How Purposeful Mindsets Drive Performance." https://www.gapinternational.com/insights/details/index/how-purposeful-mindsets-drive-performance.

Yuan, Jiahong, Wei Lai, Christopher Cieri, and Mark Liberman. 2023. "Using Forced Alignment for Phonetics Research." In *Chinese Language Resources*, edited by Chu-Ren Huang, Shu-Kai Hsieh, and Peng Jin, 289–301. Text, Speech and Language Technology 49. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-38913-9_17.

Appendix I

Tags and tiers for EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND brilliance theme

EXTREMENESS / INTENSITY / OBSESSION / ABOVE AND BEYOND	
Tier	**Tags**
Pragmatic	semantic syntactic
Rhetoric	lexical repetition syntactic repetition contrast personification escalation negation minimizing figurative expanding enumeration semantic doubling opposition certainty hyperbole pronoun shift

Word Choice	emotional charge absolute intensifier pragmatic modal verb number interjection
Hands	one hand spatial two hands spatial one hand word reinforcer two hands word reinforcer
Eye Gaze	up right up center up left down right down center down left focused gaze closed
Body	lean in lean out shoulder movement fix posture head nod affirmative head nod negative
Face	tense surprise anger happiness sad fear disgust smile narrow eyes widen eyes
Phonetic	stress breathy lower pitch raise pitch lower volume raise volume lengthening

7. Appendix II

The following table lists tags and tiers for the following brilliance themes:

- joy / excitement / exhilaration / spirit
- possibility / creation / choice / freedom
- curiosity / discovery / learning / growing
- at-stakeness / conviction / resolve / perseverance
- togetherness / relationship / connection / love
- purpose / contribution / world impact / legacy

Tags specific to certain themes are listed below the dashed line.

	JOY	POSSIBILITY	CURIOSITY	AT-STAKENESS	TOGETHERNESS	PURPOSE
	EXCITEMENT	CREATION	DISCOVERY	CONVICTION	RELATIONSHIP	CONTRIBUTION
	EXHILARATION	CHOICE	LEARNING	RESOLVE	CONNECTION	WORLD IMPACT
	SPIRIT	FREEDOM	GROWING	PERSEVERANCE	LOVE	LEGACY

Tier	Tags					
Semantic and Thematic	pragmatic universal embodied					
	magnitude joy from others for others success gratitude experiential	magnitude openness agency modal verb	magnitude joy from others for others success gratitude experiential	modal verb agency negative motivation	empathy interdependence environment mutual benefit	for others greater good commoditization knowing thyself values metaphysical
Rhetoric	figurative negation lexical repetition syntactic repetition					
	semantic repetition	contrast uncertainty	contrast semantic repetition	certainty	contrast semantic repetition	contrast semantic repetition
Eye Gaze	up right up left up center down right down left down center focused gaze closed					
Hands	one hand spatial two hands spatial one hand word reinforcer two hands word reinforcer					
Body	lean in lean out shoulder movement fix posture head affirmative head negative					
Face	tense smile narrow eyes widen eyes					
Phonetic	Pause emphatic stress pitch change rhythm					
	laughter		Laughter			
Syntactic	Conditional					
	reflexive	question	continuation question reflexive	continuation causative		causative comparative

	Absolute Intensifier emotional charge brilliance theme lexeme					
Lexical	spatio- temporal perception metaphysical agentive verb modal verb	spatio- temporal perception	spatio- temporal perception metaphysical agentive verb modal verb	spatio- temporal motion verb	agentive verb modal verb	spatio- temporal agentive verb modal verb magnitude scope

8. Appendix III

Acoustic features and their corresponding perceptual tag (used in annotation), perceptual correlate, and definition.

Acoustic Feature	**Tag(s)**	Perceptual Correlate	Definition
F0 (fundamental frequency)	**pitch change**, **emphatic stress**	Pitch	Lowest periodic cycle of the acoustic signal
Jitter	**pitch change**	Pitch irregularity	Frequency instability of F0
F1, F2, F3 (formant frequencies)	**pitch change**	Voice quality	Concentration of acoustic energy around first, second, and third formants
HNR (harmonics to noise ratio)	**pitch change**	Voice quality	Mean ratio of quasi-periodic to non-periodic signals across time segments
Intensity and amplitude	**emphatic stress**	Loudness	Measures of energy in the acoustic signal
Shimmer	**pitch change**	Loudness irregularity	Amplitude instability of F0
Speech rate	**rhythm**	Speed of speech	Number of complete elements (i.e. words / phonemes / syllables) per time unit
Articulation rate	**rhythm**	Speed of speech	Number of complete elements (i.e. words / phonemes / syllables), per time unit, excluding pauses
Pauses	**pause**	Speed of speech	Interval of silence between linguistic units

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 42

N. 1-2 – gennaio-giugno 2024

mundaneum

In copertina
Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-407-6

9 791259 654076

ISSN 1121-0095

9 770112 100950