**Literature review Summary**

1. Emotion in speech signals

[3] Speech is a non-linear time sequence signal closely related to time and emotion in the speech signal is context-sensitive information.

[9] brain is the seat of many emotions at the same time, even if at any given moment there is one dominant emotion. (blended emotions)

[9] Emotion is converyed by several multimodal cues; speech, gesture, face and physiological signals. We focus on the speech signal.

1. Emotions in HCI

[4-2] Many studies and research on affective computing increased awareness of the important role of emotion in human-computer interactions. Studying emotion theory is not only acknowledged in the HCI discipline but also expanded across many other areas of interest such as recognition and synthesis of emotion in face and body, and the influence of emotion on information processing and decision-making[4-2].

Adding emotion into computers is needed for more practical goal to function with intelligence and sensitivity towards human[4-1].

1. Availability of Emotion annotated database
2. Challenges of Emotion annotation

* [20]Emotions expressed through multiple modalities simultaneously but when annotating, the focus is usually on one single modality
* [20]The subjectivity of most annotation tasks; heavily depends on the reader’s interpretation and human usually don’t agree with each other.
* [20] increasing training improved agreement scores by Bayerl; contradicting opinion by Mohammad, over-training annotators led to confusion and apprehension in judgment tasks
* [9] main challenge addressed is categorization and annotation of real-life emotions, requiring the of a *pertinent (relevant) and limited set of labels and appropriated annotation schem*e.
* [9] *dynamic and constantly changing expression* of emotions; *context-dependent, highly person-dependent* -> unambiguous emotions are apparent in only a small portion of any real corpus; thus, the relevant emotion data are too infrequent(rare) to provide a basis for consistent annotation and modelling using fine-grained emotion labels.
* [9] major difficultiy of using matural corpora is the *expression of emotion is much more complex* than in acted speech for the above listed reasons.
* [20] (SAL/SEMAINE) labeling with everyday emotion words faces multiple problems,
  + The states occured in naturalistic data rarely fit everyday words precisely, it’s difficult to capture the rise and fall of emotion
  + Interrater (between raters) agreement tends to be low. [20-23]
  + Labeliing with dimensions have obvious attractions, and it forms the core of the scheme used here.
* [21]ML models require large manually annotated datasets. Annotation tasks range in difficulty based on the data being annotated , the annotation scheme, and the training recieved by annotators.
* [21]Most emotion detection papers use some variation of Ekman’s six core emotions or Pluttihick’s eight core emotions. Whether these categories are the best at describing human emotions is a question.
* Reasons making emotion annotation difficult [21]
  + Modality; in typical human interactions, emotions are expressed through multiple modalities simultaneously, but in annotation tasks, the focus is usually on one single modality. Emotions are expressed in speech through various means that are limited by linguistic, cultural and social constratints . In contrast, the most commonly used emotion annotation schemes are based on psychological theories that are in turn based on human interactions, not speech.
  + Thus annotating outside the originally intended modality or environment makes the emotion annotation task harder.
  + Due to subjectivity of most annotation tasks, human don’t always agree with each otehr on how to annotate’ heavily depends on the reader’s interpretation.
* [21] significant differences in the surface realization of emotion i***n different languages*** 
  + No meaning that certain emotions are not present in that language.
  + The emotion words available in different languages are s*o differen*t makes exploring emotions in speech particularly difficult.
  + C*oncep*t of a certain emotion might *only exist in one languag*e, or some emotions might be separate categories in one language and not be differentiated at all in another.
* [22] problems exhibited by existing annotation tools; namely delays between annotaion and video , laps in concentration, incaccuracy of annotation due to the sensitivity of the joystick or slider and inability to annoate remotely or online

1. [9]Representation of emotions; three types of theories generally used (appraisal/abstract/verbal categrory) ( what defines emotion)

* The a***ppraisal theory(Schere***r, 1999); providing detailed specification of appraisal(평가) dimensions, used in evaluating emotion-antecedent(이전의/ 선행) events(novelty, pleasantness, goal relevantce, etc); major problem is the only reliable way of ensuring a correct annotation is to ask the persons themselves to perform it. It accounts for individual variability in emotional reactions to the same event; peopel’s interpretations and explanations of their circumstances
* Osgood, May and Miron(1975); conceptualized several dimensions; Evaluation, Power and Activation; these are measured on scales. Evaluation scales rank emotions bad-good; Power scales strong-weak; Activity scales active-passive; other subjective dimensions including intensity, control tension etc. -> emotions are defined along ***continuous abstract dimensions*** instead of naming emotions as discrete categories. Eg. Most widely employed scheme is based on two perceptive abstract dimensions; Activation-Evaluation and has been employed to annotate several corpora with ***Feeltrace tool(Cowie*** et al. 2001)
* Most commonly verbal categories using a minimal set of emotions to be tractable(다루기 쉬운) ; most consensual set comprised of six primary basic emtions (Ekman 1992) Joy, sadness, Fear, Anger, Surprise, and Disgust.
  + Distinction between primary vs secondary or social emotions
  + ***Plutchik’s*** wheel consists of 8 primary emotions ; Ekman 6 + anticipation and acceptance
  + Secondary emotions are produced by vombinations of primary emotions that are adjacent on the emotion wheel.
  + Additional emotions can be classified as representing different degrees of intensity of primary and secondary emotions
  + Asking annotators to figure out how to assess a speaker’s internal feeling is very subjective and likely to lead to erroneous labels.

1. Single modal feature extraction vs multi-modal fusion

[3] In the process of single modal feature extraction, noise data and repetitive information have been generated which makes it harder to obtain a quality dataset, thereby affecting emotion recognition performance. [3] states that using unimodal emotion feature cannot fully describe a certain emotion of the user at the moment, whereas multimodal features allow capturing more comprehensive and detailed emotion. In addition, [3] mentioned about certain emotional information associated within and between different individual modalities, so multi-modal emotion data tries to portray the current emotion of a user from different angles which provides additional emotional information.

1. Nature of how data collected( Acted or Real Emotions?)

* [20]Lab setting/ acted dataset vs data collection in the real world
* [9] three types of corpora used; acted, induced and natural
  + - * Acted corpora; easiest to obtain and to exploit with only variations at the prosodic level since th liguistic (semantic and lexical) content has been controlled
      * Induced corpora: often obtained using Wizard of Oz(WOz) techniques and provide much more natural data. As shown in Batliner et al. (2000), the closer we get to the real-life context of interaction, the more difficult the detection of reliable emotion markers is.
      * Natural corpora; Difficult to obtain natural audio-video materials because the camera is rarely invisible; many researchers have made use of media material such as TV programmes focusing on interviews or reports that exhibit the most spontaneous situations[9]
      * Types of emotions found in natural corpora are heavily dependent on the task and situation/ context.
    - [14]In emotion recognition in non-acted speech, one often doesnt know which emotion was really intended by the speaker so the data is annotated by a group of human labelrs who don’t agree on one common class in most cases -> similar classes are systematically confused .
    - [14] From the application developer’s point of view its very important to deal with such realistic behaviour -> side effect is that one has to cope with relatively weak emotions in constast to full-blown emotions of acted speech.
    - Normally, only in a few cases, all available labelers agree on one common label -> problem of bad labeling?
* [16] many arguments in disfavour of acted emotional expression; full-blown emotions very rarely appear in the real world.
* [16] almost impossible to use natural data if basic emotions are the subject of investigation
* Problem of real emotions – make the whole setup far more complex, everybody reacts differently with respect to emotional situations.
* [16] rather relied on the performers’ ability o*f self-induction* by remembering a situation when the desired emotion had been felt strongly, known as Stanislavski method

1. Emotion models; ways to convey a specific feature of emotion

Categorical : [10] select one emotion out of a set of emotions that demotes the best feeling conveyed-> labeled

* Make use of ***Ekman’s*** six core(basic) emotions(anger, fear, happiness, sadness, surprise)
* ***Plutchik’s*** eight core emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust)
* [10] using d***omain-specific*** (domain dependent) expressive classes such as boredom, confusion;0
* most research in affective computing concentrated on six basic emotions
* However, many researchers indicate that a varied set of emotions may be requried for various fields[10-17]
* [10] pros: represent emotions wiht easy to understand emotion labels.
* [10]cons: all the emotions not included as they are grouped together by one category. Furtheremore same emotional states expressed using different emotional categories owing to cultural, environmental, linguistic or personality differences. -> lead to difficulty in fixing the actual emotional category to which it belongs.
* Thus these findings indicate that emotional categories may not represent different emotional states although the set of emotion categories is defined. -> further leads to non-optimal or incompetent emotion-detection. (limitations regarding emotion classes)
* Another problem: not possible for topics to select an appropriate class since it doesn’t occur in the label set. So categorical model has boundaries of an identiication task in attempting to distinguish the exact emotional states perceived by masses.
* However, it has been prevalent and many variation of hte model due to its ease and casualness. deviate regards how many classes they are listing;

Dimensional: [10] the representation is based on a set of quantitative measures using multidimensional scaling; (exploit rating scales for each dimension using tools like...); set of dimensions link various emotional states in this model; these dimensions respond somewhat independently resulting in dimensions that can differentially respond across time to emotion eliciting events.

* + [10] 2D (valence and arousal); Valence defines the positivit or negativity of emotion and ranges from unpleasant feelings to pleasant feeling (sense of hapiness) ; Arousal denotes the level ofexcitement that the emotion depicts, ranges from sleepiness or boredom to wild excitement.
    - [13] 2D ***Russel’s Circumplex space mod***el: valence & arousal levels organised in a circumplex shape(polar coordinate)
    - ***Thayer***[10-36] uses 2D of energy and stress
    - ***Plutchik*** [10-37] and ***Whissell***[10-38] presented model in activation-evaluation space
      * Whissell; two numericl values to show how emotions related to act-eva
      * Pluchik use angles on the emotion circle, called ‘emotion wheel’; closeness of two emotion classes in the circumplex reperesents conceptual similarity
  + 3D (valence, arousal and power); degree of powere-> as a sense of control over the emotion(influence dimension)
    - ***Mehrabian***’s model use 3D PAD(Pleasure-Arousal-Dominance) representation[10-39] (dominance; digress of feeling controlled of the state or not, Pleasure= valance)
  + [10] ***SAM***(Self-Assessment Manikin); contains pictures of manikins to evlauate the static degree of a dimension at a fixed moment (imagery based measure thought as language free; good for non english speaker)
    - [11] picture oriented questionnair developed to measure three features of emotional responses ; single-item scales that measure valence/pleasure (from positive to negative), arousal (high to low levels) , dominance/control(low to high levels)
    - [11] containing five images for each of the three affective dimensions rates on either a 9- or 21-point scale
    - Place ‘x’ on a figure or between figures (5 in total
    - Pros; what make it widely applicable: not circumscribed to any one culture and easily understood and appropriate for use in different countries ; brief, nonverval measure, used across a variety of settings with various populations , gender , age, race
    - [13] probably the most common annotating on continuous scale utilises an ordinal scale, such as SAM manikins.; argued that human estimations of relative values are more consistent than when assigning an absolute value; ranked annotations not be treated as absolute values
* Eg
* Scherer: Russel + 80 more emotion words added on
* Pros: not having to associate a certain emotional state -> can capture fine emotion concepts that differ only to a small extent
* Another pros; give a unique identification and a wide range of the emotion concepts favorable approach to measuring the distinct emotional state.
* Pros; emotional states are related to each other in a dimensioanl space, (big diff from categories)
* Pros; provide a means for measuring the degree of comparison between emotion categories; adjacent classes in the spcae are very similar, diff categories are distinctly diff form each otehr .
* Overall,dimensional emotion emodel is useful representation capuring all relevant emotions and provide a means for measureing the similarity between emotional states
* This means that each emotion can be defined as a combination of arousal and pleasantness. Not useful for distinct categories.

1. Existing Emotion annotation tools
   1. 1D vs 2D // Continuous vs per frame annotation
   2. Pros and cons of each annotation tools
   3. Similarity and differences
   4. Evaluation of the tools if possible

* FEELTRACE using joystick: drawback- delays between the annotation and the video, laps in concentration, inaccuracy of annotation due to sensitivity of joystick/ slider, inability to annotate remotely or online

JERI, , VAOAT, DANTE

* 1. [12] [10] ***Feeltrace using mouse***; studies information related to emotions continuously over time ; based on activation(how dynamic)-evaluation(how positive) space
     + Team project known as PHYSTA developed FEELTRACE to let users tract the emotional content of a speech sample as they perceive it, taking full account of gradation and variation over time



* + - Circular space, colour coding the pointer in a way users readily associate with the relevant emotional state



* + - Presenting key emotion words as landmarks at strategic points in the space
    - Developing an induction procedure to introduce observers to the system
    - 25 fitted into the space; 20 non-overlapping emotion vocabs
    - Allowing intermediate ratings , and ability to track impressions continuously
    - 2 Key issues with analysing emotion content of spontaneous dialogue; *gradation and variation over time*
      * Spontaneous speech characterised by emotional shading rather than episodes of fullblown archetypal emotion.
      * Seems tha emotional shading is rarely constant. It shifts, often gradually, but sometimes sharply, bcause of change in the speaker’s state or as it suddenly become apparent
    - Refine approach;
      * *colour-coded cursor* using a scheme derived from plutchik,- users found reasonably intuitive ; activation(red-green), evaluation(yello-blue), intermediate positions-> additive rule, white at origin
      * Add verbal landmarks of two types,
      * Representations provided by Plutchik/ and Russel provided prototypes.
      * The p*ositions of emotion word*s were set on the basis of tables published by Whissell[12-4] (activation-evalution coordinates for common emotion words)-> function to ensure users easily relate position in the space to everyday categorical descriptions of emotionality.
      * *Dimension of time represented indirectly*; bu keeping the circles associated with recent mouse positions on screen but having them shrink gradually over time. - to provide a visual indication of the way ratings were eing changed over time.
    - Inconsistent recording/ differences between raters were large enough to cause concern; some raters were intended to evoke a relatively consistent emotion, but others involved a change of mode. Results involve a clear shift from one emotional region to another.
    - System Development/ adpatation in several ways
      * Landmark reselection based on subjects’ responses to the particular representations of activation – evaluation space.
        + Users making the position in the space that they believed best represented the meaning of each word. (form users’ basic emotion vocab)
        + Data collected as part of ‘BEEVer’ procedure, designed to obtain ‘basic english emotion vocab’[12-6]
        + The aim to ensure verbal landmarks encouraged users to calibrating theri responses in the way that people on average find most natural.
      * Induction procedure developed; step by step with illustrative examples mainlyto instruct, to provide a mechanism for excluding individuals used this system in a very eccentric way (strange/weird) (BEEVer study indicated a small minority of people do seem to do that)
      * Integrated presentation; video played on a single screen; using standard pc routines for video presentations; video plaer on left, annotation circle space on the right. And
    - SYSTEM VALIDATION
      * Stimuli – clips from tv programs involving real interaction not acted
      * 16 selected, each lasted 15-30 sec.
      * Statistical analysis of resutls: show the system is reliable measurement tool and give indication of precision
        + Differences in intensity between emotional and origin (neutral stat) was highly significant -> system is capable of capturing this kind of discrimination that experimenter made in these cases
        + Also highly significant differences when comparing emotional passages ; clips differed in activation were rated differently in activation. -> means neutrality and each of four quadrants distinquished very reliably,
        + Paired t-test show significant differences between the two emo. Passages In a quadrant for three out of four quadrants. ; -> means measurements is sufficiently fine-grained resolution to discrimiate between episodes of moderately strong emotion within the same quadrant of space.
        + Overall the results indicate FEELTRACE has considerable resolving power. but critical temporal issue? (\시간상?)
        + Cannot capture distinctions betwenfear and anger as the space of possible emotions has more than 2D

1-2. Use *FEEL-trace system* to annotate powerpoint SAL data [20]

- raters annotate materials in terms of two long-established emotion dimensions, valance and activation/arousal; watch or listen to a recording and uses a cursor in adjacent window to indicate degrees of valance and arousal the individual appears to be at any given time. -> resulting in a pair of traces which show how percieved valance and arousal rise and fall as recording progresses.

- Due to important distinctions that the V-A dimensions fail to capture, SEMAINE considered a wider set of traces, each using a separarte one- dimensional scale

- Five core traces provided by every raters for each clip; valence, activation, power, anticipation, intensity

1.3 [22] O***nline tool for per-frame annotation of V-A*** based on Feeltrace using a joystick

- used to annotate the AFEW-VA dataset

- delays between the annotation and the video[22]

- lapse in concentration

- inaccuracy of annotation due to sensitivity of joystick or slider

-very few of these tool allow online remote annotations

- *addressing these drawbacks, developed an online annotation tool allowing several people to annotate video clips per-frame, for V-A remotely. [22]*

- first three addressed by per-frame annotation, and the ability to go back and forth in the sequence of correct labels. -> resulting in highly accurate annotations

- 2nd and 3rd further addressed by ability to mark the annotations in different states (to be checked again or done) and to add comments

- last inherently addressed as the develped tool is online based.

- [22] open source; easy to deploy and use, allowing multiple annotations easily and precisely; easily extended to handle more annotations e.g. discrete emotions.

2. [9] **Modified version of Transcriber;** Major, Minor(better deal with blended emotions) emotion labels wth intensity (5level scale) and control(7 level scale)

[15] **Transcriber;**  a tool for assisting in the creation of speech corpora, designed for manual segmentation and transcription of long duration broadcast news recordings, including annotation of speech turns, topics and acoustic conditions

* + Highly portable tool that can be easily installed on existing computers and environments, which works on most unix systems and windows.
  + Navigation in a long duration recording becomes an issue, as well as time-alignment of the annotations with the signal. For large quantities of data, productivity become a concern and can be increased by ergonomic tools
  + Transcribing audio or video recordings is a very time-consuming task
  + Tool should mimic as much as possible the user interfaces of standard office software, so as to reduce training time. (Easy deployment)
  + Its use should be intuitive, in order to lower the cognitive load and decrease error rates. Fast response is crucial to user actions.

3. [23] Dante using AMHUSE a multimodal dataset for HUmour SEnsing

- main aims of this study; present a novel public multimodal dataset focusing on a positive emotion; provide *DANTE, a simple, effective web-based tool for annotating emotions in the arousal and valence dimensions.*

*- dataset is the result of experiment concerning amusement elicitation using italian spoken video clips.*

*- when analysing continous dimensions of emotion, necessity of continous annotation. Challenging labelling data and laborious tasks, esp for spontaneously displayed expressions and emotions*

*- DANTE tool ; friendly and intuitive, allowing user to seamlessly annotate V-A seprately.*

- developed considering 'V/A Online Annotation tool' released with AFEW-VA datatset and 'ANNEMO'

- V-A continous dimensions

- evaluate video clip stimlus, levelof PAD; annotation made using **AffectButton**; intuitive usage; emoticon representation of emotion; no prior training is required ; the button measure emotional states in three dimensions; pleasure, arousal and dominance

- Feeltrace and Gtrace excluded bcos they rely on software intended to be installed locally and don’t allow remote annotations . .

- Aiming at web-based annotation, considered the ***‘Valence/ Arousal Online Annotation Tool’*** released with AFEW-VA dataset and ***‘ANNEMO’*** annotation tool

- ***‘Valence/ Arousal Online Annotation Tool’*** released with AFEW-V A: annotate video clips ***frame by frame*** providing continous annotations for V-A within rnage [-10 10]; tool adopted to exclusively for very short video. Average length of 50 frames, according to dataset.

- the process of *per-frame ann*otion overcomes the problems related to the **delays between the annotation and the video but introduces a bias** *resulting in sharp annotation signals that hardly follow the dynamic of expression.*

*- another drawback concerning continuous annotation tool; annoatators could have a lapse of concentration and inaccuracy due to the sensitivity of the slider.*

*-* ***ANNEMO tool :*** the one best matches the desired features; separate annotations while playing video, remote web-based framework; **limitations** *-> absense of indication for video already annnotated, or those to be done, lack of recording with fixed rate, possibility to save the annotation only on text files, missing administration interface to manage annotators and videos, unable to differentiate annotators in groups, providing different videos ->this concern decreases as the annotator training increases* -> may interfere with annotation process and the usuability of the platforma

=> these reasons why decided to develop DANTE by taking advantae of widely adopted programming languages such as PHP, Javascript, HTML, backed with a mySQL database.

- **DANTE**; allows to create a new annotator account via a dedicated admin page to assign user unique identifier for accessing personal annotation page; sidebar on the left listing all the videos assigned to the specific annotator, marked with an icon to distinguish between the already annotated videos and those to be done; the centre of the page dedicated to annotation itself, showing video and a sliding bar ranging from –1 to 1, included SAM visualization specific for selected affective dimension (V or A) to help annotators.

- validation available !!

4. [24] Continuous, real-time, joystick-based emotion annotation framework

- told subjects to indicate their instantaneous emotional state in a V-A space, using a joystick

- analysis ; SUS system usability scale questionnaire unveiled the excellent usability, MANOVA analysis of mean V-A ratings, trajectory similarity of the annotations confirmed successful elicitation of emotions, change point analysis of annotations revealed a direct mapping between emotional events and annotations thus enabling automatic detection of emotionally salient (현저한) points in the video,

- rating patterns were cohesive across participants so confirms the value, validity and usability of our annotation framework

- first introduced in 2014, similar to DARMA,

- most joysticks incorporates a return spring and hence automatically realign to the centre under no force. User consciously aware of the joystick’s position in use, reductin the cognitive load

- 2D setup of framework allow users to provide a *more comprehensive description of their emotional experiences* thatn is possibly by annotating each dimension separately.

- the annotation UI embedded in a video ; V and A axes of UI included SAM that serve as visual guides to user for determining V-A levels

- validation analysis avialbe in the lit [24]; ststistically coherent results obtained

- use of joystick; allows for continous and providing proprioceptive feedback to the annotaor ; more ergonomic than computer mice; simplicity and high acceptance of the proposed framework

- drawbacks ; abeit small, cognitive load imposed on users while annotating. Video stimuli must be carefully selected in order ot elicit the desired affect state. Given ironical sexual content, this reaction delays.

5. ANNEMO [25] annoating emotional and social behaviours

- from introducing the RECOLA multimodal corpus of Remote collaborative and affective interactions

- tested different annotation tools; Feeltrace and Gtrace; but wanted to use a web-based approach for facilitating remote annotation of data so developed own.

- Feeltrace; judging two emotional dimensions at the same time, may be too cognitively demanding to reach a high quality on both

- rather used a setting wiht one time-continous annotation for each affective dimension like GTrace.

- annotator logged in a web based annotation interface by using a unique identifier through G*oogle Chrome* which enable handling video codecs; interface split into two; left side -> scrolling list of the audiovisual recordings whereas the video and annotation cursor desplayed one below the other ; two affective dimensions annotated separately and time-continuosly using a slider with value ranging –1 to 1 with a step of 0.01

- to a*void delay in data transmissio*n, timestamps estimated on local machine as the detaly between a slider event and the video start

***- Post processin***g of annotations to reduce unwanted variabiliites in the data using different normalization techniques to study their influence on the inter-rater agreement

* + - Deal with missing values ; *interpolated dat*a using *piecewise cubic interpolation* -. it preserves monotonicity and shape of data
    - Then binned data with a frame rate fixed to match with the one used in video recording. -> reduce jump effects due to reannotation
    - Varaiablilites in human judgements do exist, use a local normalization for each sequence for each annottator; *zero mean(ZM)* to remove an eventual bias; s*ynchronization* to tackle the issue of having differnet time reactions betwen annotators; synch delay of a give annotation estimated by *minimizing the inter-rater mean squared error (MSE)* pair wisely with the annotations provided by all other annotators.
  + ANalysis results available!!
  + Anaylsis of annotation show a good inter-annotator ageement rate for affective dimensions. But relevant of this tech for emotion recognition is questionable as the balance of instances can change significantly.

6. GTrace

- allows users to create a 'trace' that specifies how they *see emotion rising and falling over tim*e. They respond by using *a mouse* to move a cursor in a 1-D window which appears beside the material to be rated. Users can select from a range of pre-specified scales, or form their own

- Feeltrace and related programs widely used but suffers from two major limitations; first, code is over a decade old, and only runs on Windows XP, Second, oriented to tracing in a circular space with two dimension, V and Activation. -> although clearly useful, it’s also cleaar that they are not the only dimensions that one might want to trace.

; 2nd generation systems that allows users to trace a substantial number of dimensions in 1d windows but yet not been freely available

* + Gtrace ; a Feeltrace successor system providing *multiple 1D traces* on current windows systems, and o*utput compatible* with th*e W3C standard EmotionML[26-4] (a new standard describing emotion on the web)*
  + *Devices to help raters to understand the scale in the way they were intended; -* ***colour of the cursor(fo****rm acoloured disc) changes as it moves along the scale, in a way that goes naturally with the meaning of the scale. Eg. Pure red at the negative extreme, pure green at the positive extreme*
  + - leaves a *tail behind* it, in the form of circles that show where it was recently, and **s*hrink away over t****im*e.
  + Folders containing scales based on different category terms ; *‘Big 6”* ; basic emotions proposed by Ekman, “Everyday” dimentional descriptions ‘***FSRE’ scales by Fontaine et al [26-6] -***> shows four dimensions (valence, power, arousal/activation, anticipation(less common but reasonabley intuitive));
  + Other folder contains a scale dealing one other dimension, intensity. Also available to add other scales derived from the EmotionML vocabularies.
  + Reliability of these dimension contained in [26-2/3]
  + At start, ask user to select one of vocab available in EmotionML. Tracing scale displayed on the right ; can select video clips from media folder and scale panel to list scales aviable to use in folders
  + System is desinged to allow previews and multiple attempts
  + Xml output provided for EmotionML scales
  + *Download zipped file from website not available.*
  + *Must first install Microsoft .Net Framework version 4.*
  + *Minimum screen resolution of 1280 x 1024 required*
  + Trace-type techniques are as general as questionnaries ;
  + But The fact that the code is available means that it can be adapted for applications that the current form doesn’t suit; ex, using touch screen interfaces or 2d rating like feeltrace.

6. CARMA [27]

- collects continous ratings while displaying audio and video files

- desinged to be highly user-friendly and easily customizable

- based on Gottaman and Lvenson’s affect rating dial

- provide moment-by-moment ratings of multimedia files using a computer mouse or keyboard

- continuous measurement system, EmuJoy, FeelTrace, Gtrace are more focused on continuous rating. But, most fallen out of repair, and are difficult for non progrmmers to implement and cutomize.

= EmuJoy and FeelTrace require raters to respond on a 2D scale which is difficult for them and increases cognitive load of the tasks

-> **Carma** fills the need by providing a focused media annotation solution that can be easily customized and used by researchers/ participants on a variety of multimedia types.

* Carma, Written in MATLAB programming language and compiled into Windows application using the matlab compiler. Executed using free matlab compiler runtime software from mathworks, media playback is implemented using the Windows Media Player ActiveX controller-> can load any multimedia file formats compatible with the version of windows media player installed on the computer.
* Using slider to rate the selected media file in real-time as it plays.
* Adj to slider is a customizable rating scale that provide labels; users can also customize the gradient displayed on hte scale by selecting any three colours.
* Setting to customize scale for dimension, at the end, annotation viewer window opens to display collected rating, -> user can playback the media file and view ratings or export ratings
* Explort annotation files as microsoft excel spreadsheet containing mean ratings for each sec, corresponding timestamps and metadata of Carma configuration when annotated
* Only available in windows due to dependency on ActiveX, Carma does not run on Linux or Mac
* Dependencies on MATLAB Compiler Runtime 8.3 32bit Windows version and Windows Media Player via ActiveX
* Previous studies focus on broad affective dimensions such as V, A, power. But due to it***s customizable rating scal***e, Carma can be adapted to nearly any project for which continous ratings are desired. Specific emotional and cognitive states can be used in place of affective dimensions, allowing raters to report on experiences of anxiet, confusion pain, etc. It can even be used to annotate non-affective information such as quality of teaching in video of a lercture.
* Measurements also be used as training data for automated systems attempting to detect and anlayse affective states from audio and video. Collecting responses from a variety of raters is critical for such a task, *and easy to use programme* like carma can greatly *facilitate this data collectio*n.
* Interpid (intermed) users may *rewrite the program* in open source MATLAB alternatives, or adapt the system to co*llect ratings over the web*.
* May add self-report questionnaires or experimental manipulations befor or after collection of ratings by adding additional windows to the baseline programe,
* The spreadsheet format of carma’s output is ideally suited to incorporating additional data fields.

7. EMuJoy [28] software for continous measurement of perceived emotions in music

- Emotion measurement with Music by using a joystick

* Enhances Schubert’ 2DES software in three ways;
  + - 1; give self-reports for different media in real time with mouse , joystick or other medium. Use of joystick advantageous
    - 2; numerous useful features ; completely controlled over internet
    - Ex. Any audiovisual stimuli presented to subjects without any personal contact with researcher. Even data acquisition take place over the internet
    - 3; platfomr independent solution; also open to modification such as customization for further multimodal stimuli
  + Chose a 2D emotion spacee with dimensions V-A; proposed to measure both V and A on bipolar scales ; neither of pleasure-displeasure/ sad –happy to elicit V ratings in order to avoid a bias resulting from pre-conceived terms
  + Further developed interface from 2DES interface; using a pointer to depict the position iwthin emotion space. Controlled with a joystick or computer mouse.
  + Contrast to 2DES, whole screen used for emotion space, media file played at the background, embedded in the background
  + Change in colour of the pointer; resemble the shape of a worm with a tail and a face, which makes the handling very intuitive, and uncomplicated
  + Looking at the face, the user gets intuitive feedback within the emotion space on their self-report. ; tail traverses the tracetory of previous points in the space to illustrate the dynamics of responses. This related to that of IAPS, also using f*acial icons* for its SAM.
  + Communication between server and local application is based on the TCP protocol
  + Bcoz of open source nature of the software, no limits in how it could be further developed and adapted for any needs.

8. DARMA ; [29] software for dual axis rating and media annotation

- modernized continous measuremtne system that synchronizes media playback and the continuous recording of 2D measurements; these measurements can be observational and self-reported and provided in real-tme through the manipulation of a computer joystick.

- also prov*ide tools for reviewsing* and comparing collected measurements and for customizing various settings.

- domain- independent software tool , open-source using the GNU General Public License (GPL) and downloadable from ....

- provid*e moment-by-moment rating*s of multimedia files

- configured on a number of parameters including its labels and numerical range

- continuous measurement systems rely on raters to provide high frequency reports of their perceptions or experiences. Rather than repeatedly stopping raters to collect theri reports, measurements are unobtrusively sampled from an apparatus that can be continuously manipulated by the rater.

For ex, raters may be trained to control joystick..... Data about the current position of apparatus can be collected at a high frequency and saved for later use. As a result, dimensional measurements can be captured in real time.

- One continuous measurement software that bears highlighting is ***CARMA,*** which included a n*ovel suite of features for reviewing collected measurements. These features increase the efficiency of traning and quality control.* Carma needs measurements to be made on one dimension at a time, but, many research areas are primarily interested in collecting 2D measurements.

- 2D measurements often derive from circumplex models, variables arranged into a circular space defined by two continuous (bipolar) dimensions. The i*ntersection of 2D mesaurements allows for much richer description than either dimension could provide alone*, as they provide coverage of interstitial space defined by blends of two dimensions.

- two popular circumplex models ; affective circumplex (Russell) describe emotion using V-A dimensions; interpersonal circumplex (Horowitz et al. ) describing social interaction using dimensions of agency(submission to dominanace) and communion(coldness to warmth)

- increasingly common in clinical and personality psychology to collect ratings of social communication using the interpersonal circumplex.

\*\* As such, **a feature-rich software package** for collecting 2D/circumplex rating is needed.

* + Darma; open-source software package that builds upon CARMA.
  + Used to collect observational measurements of behaviour or self-report of various aspects of experience.

- customized to collect *2d measurements of any kind*

- Darma written in *MATLAB programming languag*e , compiled into a standalone application for microsoft windows using matlab compiler. it has two software dependencies. 1; open source VLC Media Player, 2. freely available MATLAB runtime to enable execution of compiled matlab applicaitons

* 1. - in order to collect ratings on 2D simulataneously it requires users to maniputlate a computer joystick. Greater precision of control offered.
  2. *2 windows ; Collect Ratings window and* ***Review Ratings window***

*- review ratings window enables users to view previously saved annotation files alongside the original media file. Annotation files from any raters can be loaded simulteneously. Options for visulaizing each rater’s measuremntes and for assessing inter-rater agreement and reliability. Mean of multipe rater’s measuremetns, mean series, Analyze Ratings options for desciptive stats. Etc.*

*- Darma 1st software implementation of conti, measurement system that combines playback-collection synchronization, qualitative and quantitative review tools, and ability to collect 2D measurements. -> these increase efficiency and quality of important research task currently being tackled with old tech. Automatic synchronization, pause & resume whenever necessary allows longer media segments to be measured at one time, prevent data from being lost due to unplanned interruptions. Annotation review is powerful tool for observer training and quality control.*

*- EVALUATION for usuability of software avialble !! User SURVEY(satisfaction) also present!*

*- in response to the user satisfaction survey, future updates will focus on improving clarity of the installation process and making documentation and technical support more accessible to user; add more tools to stramline the analysis and management of the annotation files.*

*- essential features of darma set it a step above existing 2D systems such as playback-collection synchronization, annotation review tools and painless customization.*

*- Continuos measuremnets systems talble present!!!!*

- ------------------------------------------------------------------------------------------------------------------------

* [21] recent survey studies show most modern research on emotions, particularly in NLP, is at least to some extent based on the work of Ekman, Plutchik, esp his Wheel of Emotions and ***SenticNet[21-11](***modified Plutchik’s wheel to show change in emotional intensity on Gaussian curve)
* [21] work of Keltner and Cowen tried to tackle categorization of emotions in a number of studies. 27 distinct emotion categories by studying emotion responses to a numver of different stimuli such as video, music, facial expressions, speech.

-> emotion annotation scheme partly relied on in ***GoEmotions [21-16]***

1. Joysticks or mouse based tool comparisons [avialable tools for anotation] [24]

[24] most approaches still using mouse based tool; user continuously press a button during annotation, -> inherent cognitive and physical loads associated with continous annotation, problem more evident in using mouse based tools

* More comprehensive reporting of emotional experience allowed by simultaneous annotation of VA in 2D space -> but often not pursed due to lack of tools or due to concerns regarding congnitive overloading.
* To overcome these two issues, using joysticks, Joysticks more intuitive to use than mouse based tools, and 2D tools that allow for simultaneous acquisition of V and A ratings have become increasingly prevalent.
* This approach in their early stage of devlopement, so still most current joystick based implementations acquire only 1D from each user. and most still using mouse based tools

1. Preferred Platform type for user these days
2. Improvement in existing design

* [9] in emotion recognition from natural speech, it’s unenvisagable (unpredictable) to have perfect emotion reference labels. So as proposed in Steidl et al. 2005 the machine learning results must be compared with the performance of human labellers. To evaluate the classification results which takes into account systematic labelling errors

1. Plan for system validation (Survey)

Further work includes evaluation of .... based model proposed searlier to incorporate dimentsions as well.

Our contribution is to incorporate the space into a system that observers find it easy to use, and that gives reasonably reliable outcomes

Incorporate several features designed to convey the idea of emotion as a point in a 2 D spacel

A web site has been designed for the distribution of the tool

Why existing annotation tools could not fulfil our needs

Investigation of emotional cues in speech is gaining growing attention. Due to the new developments with respect to human-machine interfaces that see applications of automatic recognition and simulation of emotional expression within reach.

One of the biggest hurdles for utilization of ml in interdisciplinary projects is the need for annotated training data which is costly to create[21].

[21]Emotion annotation is notoriously difficult task, the current annotation scheme based on psychological theories of human interaction are not always the most conductive for the creation of reliable emotion annotations. Nor are they optimal for annotating emotions.

Emotion labels are usually obtained via either manual annotation which is tedious and time-consuming, or questionnaires, which neglect the time-varying nature of emotions and depends on human’s unreliable introspectiion. To overcome these limitations, we developed a continous, real-time, joystick-based emotion annotation framework [24]. Manually annotated using discrete emotion labels method unsuitable when using dynamic stimuli(video) as not considering time varying nature of emotions. Discrete emotion labels are insufficient to define strength of emotional experiences. Thus*, RECENTLY* researchers started suing annotation tool taht allow for *continuous reporting*. These mostly based on 2d circumplex model of emotion by Russell, wherein emotional labels are represented in continous dimensions of V and A. commonly available 1D and 2D tools ; CMS[24-9], GTrace, CARMA, FEELTRACE, EMUJoy, DARMA.

[24] by robustly estimating emotions in real-time, machines can imporve interaction experiences for human. Modify its behaviour to ... in human-robot interaction. The data from pysiological sensors on a device need to be linked to the internal emotions experienced by human, and this association step is still a largely unsolved problem. Emtoion –inducing stimuli (videos, music or photos) provided to participants and measure theri affective response using biosignals, speech signals or computer vision based approaches.

[24] Emotional memory is short-termed ; highly desirable that the participants report/ annotatae theri affect state asap, ideally continuously while emotional stimuli are presented. But annotation exercise must be as unobstructive as possible, such taht it doesn’t influence the emotions of participant annotating her emotional experience.

[23] Issue of gathering good affective data is crucial for learning models and subsequent benchmarking. Early research on affective compurting was mostly focused on facial expression recognition, and datasets built on a single modality, namely images or video.

*Annotation reliability* is usually measured by calculating inter-annotator agreement, referred to as inter-rater correlation.

Bayerl et al.[21-6] found that increasing training improved agreement scores ,

Mohammad [21-29] found that minimal guidelines improved agreement scores because over-training annotators led to confusion and apprehension in judgement tasks.

[22] Co*ntinours dimensional models* of human emotion, based on V-A, shown to be more accurate in describing a broad range of spontaneous, everyday emotions than the more traditional models of discrete stereotypical emotion categories.

But most prior work on estimating V-A considered only ***lab setting and acted data***

[22] proposes a new dataset of highly accurate per-frame annotations of V-A for 600 video clips; per frame annotations of 68 facial landmarks

-> results; method that perform well on constrained datat dont necessarily generalise to uncontrolled data and vice-versa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |