

# La sentiment analysis come strumento di studio del parlato emozionale?

**Paolo Mairano**  
University of Lille, France  
paolo.mairano@  
univ-lille.fr

**Enrico Zovato**  
University of Turin, Italy  
ezovato@  
gmail.com

**Vito Quinci**  
University of Turin, Italy  
vito.quinci540@  
edu.unito.it

## Abstract

**Italiano.** Vari studi in letteratura hanno dimostrato che il parlato emozionale è caratterizzato da vari indici acustici. Tuttavia, tali studi hanno quasi sempre utilizzato parlato recitato, ignorando il parlato elicitato in maniera ecologica a causa della difficoltà nel reperire adeguate produzioni emozionali. In questo contributo, esploriamo la possibilità di utilizzare la *sentiment analysis* per selezionare produzioni emozionali da corpora orali. Abbiamo utilizzato il corpus *LibriSpeech*, da cui abbiamo estratto valori di *sentiment analysis* a livello di frase e di parola, nonché vari indici acustici e spettrali associati al parlato emozionale. L'analisi della relazione tra i livelli acustico e testuale ha rivelato effetti significativi ma di portata ridotta. Questo ci fa pensare che tali due livelli (acustico e lessicale) tendano a essere relativamente indipendenti, rendendo inappropriato l'utilizzo di metriche testuali per la selezione di materiale acusticamente emozionale.

**English.** *Abundant literature has shown that emotional speech is characterized by various acoustic cues. However, most studies focused on sentences produced by actors, disregarding ecologically elicited speech due to difficulties in finding suitable emotional data. In this contribution we explore the possibility of using sentiment analysis for the selection of emotional chunks from speech corpora. We used the LibriSpeech corpus and extracted sentiment analysis scores at word and sentence levels, as well as several acoustic and spectral parameters of emotional voice. The analysis of the relation between textual and acoustic indices revealed significant but small effects. This suggests that these two levels tend to be fairly independent, making it improper to use sentiment analysis for the selection of acoustically emotional speech.*

## 1 Introduzione

L'espressione delle emozioni può avvenire attraverso diversi componenti a vari livelli linguistici (Reilly & Seibert, 2003): lessicale (verbi modali, elementi rafforzativi, attenuativi, o valutativi), sintattico (es. le proposizioni relative possono commentare azioni e comportamenti), acustico (prosodia, qualità della voce), e paralinguistico (espressioni del viso, gesti). I framework tradizionali per l'analisi delle emozioni sono basati su categorie (Ekman, 2000) o su dimensioni (Russell, 1980). I primi distinguono vari stati emozionali (rabbia, gioia, paura, tristezza, etc.), mentre i secondi tendono a definire le emozioni come coordinate in uno spazio multidimensionale, in cui ogni dimensione rappresenta una proprietà di uno stato emozionale. Tra i numerosi framework esistenti, Russell (1980) ipotizza due dimensioni: *valence* (valenza, positiva vs. negativa) e *arousal* (attivazione, alta vs. bassa). La classificazione degli stati emozionali tramite indizi linguistici si è rivelata un compito arduo tanto nei framework categoriali quanto in quelli dimensionali, e l'interazione dei vari livelli linguistici complica ulteriormente la situazione: non è ancora chiaro se la componente lessicale / sintattica debba essere considerata come dipendente o complementare alla componente acustica.

Nonostante tali problemi, molti studi hanno analizzato il parlato emozionale con l'obiettivo di individuare i correlati acustici specifici dei vari stati emozionali. Alcuni studi hanno dimostrato che variazioni sistematiche della frequenza fondamentale (sia in termini di *pitch range*, sia in termini di *pitch medio*) accompagnano realizzazioni di parlato con valenza positiva (Burkhardt & Sendlmeier, 2000). Ma anche altri parametri prosodici sembrano avere un ruolo importante nella comunicazione delle emozioni: sono stati

riscontrati effetti dell'intensità e della velocità d'eloquio (Johnstone & Scherer, 2000); infatti, varie misure acustiche (deviazione standard della frequenza fondamentale, energia media, durata dei periodi, *spectral-dropoff*, etc.) sono state usate per predire i giudizi di parlanti madrelingua (Banse & Scherer, 1996) e vari altri parametri sono stati usati in altri studi (cf. Schröder et al., 2001, e Audibert, Aubergé & Rilliard, 2005).

Tuttavia, uno dei limiti di questi studi riguarda l'affidabilità dei dati: data la difficoltà di elicitarne parlato emozionale controllato, gran parte degli studi utilizza registrazioni di parlato recitato, che spesso risulta stereotipato o esagerato (Scherer, 2003). In questo contributo, verifichiamo se la *sentiment analysis* (d'ora in poi: SA) possa essere d'aiuto in questo senso. La SA, ovvero lo studio delle opinioni, sentimenti, recensioni delle persone in forma testuale (Liu, 2003), è un settore NLP in rapida crescita, grazie anche all'ampio ventaglio di applicazioni, quali la classificazione di email (Mohammad & Yang, 2011), romanzi (Mohammad, 2011), recensioni cinematografiche (Sadikov, 2009), recensioni di articoli o servizi acquistati (McGlohon, Glance & Reiter, 2010). I sistemi di SA vanno da metodi a regole relativamente semplici, fino a tecniche avanzate di *deep learning* - vedi Liu (2012) per una rassegna.

In questo studio, verifichiamo la relazione tra i valori di SA e le caratteristiche acustiche del parlato letto elicitato in maniera naturale, estrapolate da audiolibri. Il fine ultimo è quello di estendere l'analisi a dati di parlato spontaneo; tuttavia, dati i numerosi problemi che questo tipo di parlato comporta, abbiamo preferito iniziare da dati di parlato letto in cui le emozioni non fossero state elicitate esplicitamente. Per misurare il grado di emozione espresso dal testo degli audiolibri, sono stati utilizzati *SentiWordNet* (Baccianella, Esuli & Sebastiani, 2010) e *Vader* (Gilbert & Hutto, 2014), che operano principalmente a livello lessicale. Sul piano acustico, abbiamo estratto vari indici (per lo più prosodici) descritti in letteratura. Un'analisi simile a questa, che studia la correlazione tra SA e parametri acustici, è stata condotta da Charfuelan & Schröder (2012) su dati di un solo speaker e di un solo audiolibro. Qui estendiamo l'analisi a 251 audiolibri letti da speaker diversi, nella speranza che i risultati abbiano sia rilevanza teorica (studio dell'interazione tra livello lessicale e acustico nel parlato emozionale), sia un risvolto pratico

(utilizzo della SA per la selezione di parlato emozionale non recitato).

## 2 Dati e metodologia

### 2.1 Corpus

Per studiare la correlazione tra i valori della SA e le caratteristiche acustiche del parlato emozionale, abbiamo utilizzato *LibriSpeech* (Panayotov et al., 2015), un corpus open-source contenente circa 1000 ore di parlato in inglese. I dati di *LibriSpeech* provengono a loro volta dal progetto *LibriVox* (una collezione di audiolibri di dominio pubblico, disponibili su [librivox.org](http://librivox.org)), e i testi sono stati segmentati e allineati automaticamente dagli autori del corpus. Ai fini di questo studio, abbiamo limitato l'analisi alla sezione *train-clean-100* del corpus (contenente 100 ore di parlato corretto e pulito, originariamente concepito come *training set* per sistemi ASR), che include i dati di 251 audiolibri. I lettori sono un mix di professionisti e non-professionisti di sesso maschile e femminile (l'età non è riportata). L'elenco dei testi registrati (consultabile sul sito web di LibriVox) include principalmente opere letterarie britanniche e americane, antiche e moderne.

Tutto il materiale è stato trascritto foneticamente con il front-end del sistema TTS *Vocalizer* di *Nuance Communications*, secondo il modello di General American. Le trascrizioni sono poi state allineate al segnale acustico e infine convertite in formato *TextGrid* per essere utilizzate con *Praat* (Boersma & Weenink, 2018).

### 2.2 Metriche di *sentiment analysis*

I valori di SA sono stati estratti dal testo di ciascuna frase usando strumenti open-source, quali *Vader* (Gilbert & Hutto, 2014) e *SentiWordNet* (Baccianella et al., 2010), entrambi disponibili nella libreria NLTK di Python. Si tratta di strumenti classici nella letteratura sulla SA e relativamente semplici dal punto di vista dell'utilizzo e dell'implementazione (trattandosi di sistemi a regole). In futuro, l'analisi potrebbe essere estesa utilizzando strumenti più complessi e sofisticati, come i moduli di SA dei progetti *OpENER* (<http://www.opener-project.eu/>) e *StanfordNLP* (<https://nlp.stanford.edu/>).

*Vader* fornisce tre valori: (a) un punteggio di polarità positiva compreso tra 0 e 1 (*Vader\_comp*), (b) un punteggio di polarità negativa compreso tra 0 e 1, (c) un punteggio

derivato dagli altri due compreso tra -1 (negativo) e +1 (positivo). Questi valori sono ricavati grazie a un sistema a regole, basato sul lessico *Vader*, nel quale le parole sono associate a un punteggio di polarità ottenuto dalle valutazioni di 13 madrelingua. *SentiWordNet* adotta un approccio leggermente diverso: le parole nel suo lessico sono associate a punteggi di polarità positiva o negativa procurati tramite un'analisi quantitativa di ogni synset (vedi Baccianella et al., 2010, per maggiori dettagli).

I valori di *Vader\_comp* sono stati valutati sulla base di un sottoinsieme di 1000 frasi annotate manualmente da uno degli autori (prendendo frasi isolate, quindi senza informazioni sul contesto o sul co-testo), ottenendo un'accuratezza pari al 72%.

### 2.3 Indici acustici del parlato emozionale

Sebbene la maggior parte degli studi si concentrino sui parametri acustici a livello di frase, noi abbiamo applicato l'analisi anche a livello di parola, sulla base dell'ipotesi che le parole con carica emozionale possano essere caratterizzate da specifici indici acustici (Tsiakoulis et al., 2016).

Per l'analisi a livello di frase, gli indici acustici sono stati estratti per ogni frase. Per l'analisi a livello di parola, invece, gli indici acustici sono stati estratti dalla vocale accentata delle parole non funzionali al fine di controllare le differenze spettrali dei vari fonemi vocalici (il fonema vocalico è stato incluso come fattore nell'analisi statistica). I seguenti indici acustici sono stati estratti tramite *Praat*: F0 mean (frequenza fondamentale media in semitoni), F0 stdev (in semitoni), F0 range (0.05-0.95), F0 max (0.95), F0 min (0.05), shimmer, jitter, Hammarberg index (HAM, differenza tra il massimo di energia nelle bande 0-2 kHz e 2-5kHz, cf. Hammarberg et al., 1980), Do1000 (riduzione di energia spettrale oltre 1000 Hz), Pe1000 (energia relativa a frequenze oltre 1000 Hz vs energia sotto i 1000 Hz, cf. Scherer, 1989, e Drioli et al., 2003). I valori di F0 sono stati estratti tramite il metodo di autocorrelazione di *Praat* (con i parametri di default) secondo una procedura in 2 fasi: in una prima fase, l'estrazione è stata fatta con un range fisso 75-400 Hz; l'intervallo interquartile (IQR) è stato calcolato sui valori così ottenuti, e una seconda estrazione è stata realizzata nel range tra +50% e -25% dall'IQR.

Inoltre, per l'analisi a livello di frase abbiamo estratto la durata totale in ms dal primo

all'ultimo fonema (DUR), *speech rate* (SR, numero di fonemi diviso la durata complessiva incluse le pause), *articulation rate* (AR, senza le pause), *pause/speech ratio* (PSR).

Tutti i parametri acustici estratti sono stati trasformati in *z-scores* per ogni speaker, nel tentativo di normalizzare le differenze tra speakers. Le frasi contenenti meno di 3 secondi di parlato sono state escluse dall'analisi. Per ogni parametro acustico, i valori che si scostavano >2.5 deviazioni standard dalla media sono stati esclusi come probabili errori di detezione.

## 3 Risultati

### 3.1 Analisi a livello di frase

I dati sono stati analizzati su *R* tramite modelli a effetti misti con la libreria *lme4* (Bates et al., 2014) per valutare la relazione tra valori di SA e parametri acustici. In una prima analisi, abbiamo costruito dei modelli per valutare l'effetto di *Vader\_comp* (che prendiamo come indicativo di valenza) su ogni indice acustico separatamente, includendo sempre il fattore speaker come effetto aleatorio, es.:  $F0\_range \sim Vader\_comp + (1 | speaker)$ . Questa prima analisi ha rivelato che il valore di *Vader\_comp* ha un effetto significativo sui valori di F0, in particolare F0 max, F0 range, F0 mean e F0 stdev (v. tabella 1).

Modello	p val.
$F0\_min \sim Vader\_comp + (1   speaker)$	ns
$F0\_max \sim Vader\_comp + (1   speaker)$	***
$F0\_range \sim Vader\_comp + (1   speaker)$	***
$F0\_mean \sim Vader\_comp + (1   speaker)$	***
$F0\_stdev \sim Vader\_comp + (1   speaker)$	***
$AR \sim Vader\_comp + (1   speaker)$	ns
$PSR \sim Vader\_comp + (1   speaker)$	$p = .05$
$Shimmer \sim Vader\_comp + (1   speaker)$	ns
$Jitter \sim Vader\_comp + (1   speaker)$	ns
$HNR \sim Vader\_comp + (1   speaker)$	ns
$Do1000 \sim Vader\_comp + (1   speaker)$	ns
$Pe1000 \sim Vader\_comp + (1   speaker)$	ns
$HAM \sim Vader\_comp + (1   speaker)$	ns

**Tabella 1.** Effetto di *Vader\_comp* sui valori acustici.

L'effetto di *Vader\_comp* non è risultato significativo per la predizione degli indici di ritmo e durata. Quindi abbiamo voluto verificare se questi parametri si correlino con l'intensità di attivazione, piuttosto che con la valenza. Abbiamo quindi valutato modelli separati per frasi negative (*Vader\_comp* < 0) vs positive

( $Vader\_comp > 0$ ). Tali modelli hanno mostrato che il valore  $Vader$  di positività (range:0-1) ha un effetto significativo non solo sugli indici di F0, ma anche su AR, PSR, shimmer, HNR, Do1000, Pe1000 e HAM. Analogamente, il valore  $Vader$  di negatività (range:0-1) ha un effetto significativo per gli indici di F0, nonché su AR e shimmer (v. tabella 2).

Effetto	Vader>0	Vader<0
<i>F0 min</i>	***	***
<i>F0 max</i>	<i>ns</i>	***
<i>F0 range</i>	*	**
<i>F0 mean</i>	**	<i>ns</i>
<i>F0 stdev</i>	*	*
<i>AR</i>	***	***
<i>PSR</i>	*	<i>ns</i>
<i>Shimmer</i>	***	**
<i>Jitter</i>	<i>ns</i>	<i>ns</i>
<i>HNR</i>	*	<i>ns</i>
<i>Do1000</i>	*	<i>ns</i>
<i>Pe1000</i>	*	<i>ns</i>
<i>HAM</i>	*	<i>ns</i>

**Tabella 2.** Effetto di  $Vader\_pos$  e  $Vader\_neg$  sui valori acustici.

Questi risultati sembrano quindi suggerire che gli indici di F0 siano influenzati dalla valenza della frase, mentre gli indici ritmici e spettrali si correlano con l'intensità di positività o negatività della frase. Tuttavia, la parte di varianza spiegata dai vari modelli rimane bassa, con ad esempio  $R^2 = 0.01$  per il modello che predice AR.

Infine, abbiamo costruito un modello a effetti misti per predire  $Vader\_comp$  a partire dagli indici acustici, includendo il fattore 'speaker' come effetto aleatorio. Dopo l'eliminazione degli effetti non significativi, abbiamo ottenuto  $R^2 = 0.06$  per la contribuzione cumulativa di tutti gli indici acustici significativi. Considerando separatamente le frasi con valori positive e negativi (cercando quindi di predire i valori  $Vader$  di positività e negatività sulla base degli indici acustici),  $R^2$  sale a 0.09 per il modello che predice i valori  $Vader$  di positività, e a 0.12 per il modello che predice i valori  $Vader$  di negatività.

### 3.2 Analisi a livello di parola

Analogamente a quanto fatto a livello di parola, in una prima analisi abbiamo costruito dei modelli a effetti misti per valutare la relazione tra valori di SA e ognuno dei parametri acustici separatamente. Come variabile *predictor* abbiamo utilizzato il valore di valenza per ogni

parola nel lessico di  $Vader$ , e abbiamo incluso il fattore 'speaker' come effetto aleatorio. Inoltre, per i parametri spettrali HNR, Do1000, Pe1000 e HAM abbiamo incluso il fattore 'fonema' come effetto aleatorio, poiché tali parametri variano in funzione delle diverse vocali. Come per l'analisi a livello di frase, i modelli ci dicono che il valore di valenza di  $Vader$  ha un effetto significativo sugli indici F0 min, F0 range, F0 mean, F0 stdev, e questa volta anche shimmer e jitter (v. tabella 3).

Modello	p val.
<i>F0 min ~ Vader + (I speaker)</i>	***
<i>F0 max ~ Vader + (I speaker)</i>	<i>ns</i>
<i>F0 range ~ Vader + (I speaker)</i>	***
<i>F0 mean ~ Vader + (I speaker)</i>	***
<i>F0 stdev ~ Vader + (I speaker)</i>	***
<i>Shimmer ~ Vader + (I speaker)</i>	***
<i>Jitter ~ Vader + (I speaker)</i>	**
<i>HNR ~ Vader + (I speaker)</i>	<i>ns</i>
<i>Do1000 ~ Vader + (I speaker)</i>	<i>ns</i>
<i>Pe1000 ~ Vader + (I speaker)</i>	<i>ns</i>
<i>HAM ~ Vader + (I speaker)</i>	<i>ns</i>

**Tabella 3.** Effetto di  $Vader$  sui valori acustici.

In una seconda analisi, come a livello di frase, abbiamo voluto verificare se i parametri acustici fossero correlati all'intensità di attivazione positiva o negativa della parola. Per far questo, abbiamo costruito altri modelli separati per parole con valenza positiva (*SentiWordNet pos value* > 0) in frasi positive ( $Vader\_comp > 0$ ) e per parole con valenza negativa (*SentiWordNet neg value* > 0) in frasi negative ( $Vader\_com < 0$ ). I modelli relativi a parole positive hanno rivelato un effetto significativo di *SentiWordNet pos value* su HNR, Do1000 e Pe1000, ma solo marginalmente significativi sugli indici di F0. I modelli relativi a parole negative in frasi negative hanno rivelato un effetto significativo di *SentiWordNet neg value* su HNR, Do1000, Pe1000, e HAM (v. tabella 4).

Nell'analisi a livello di frase, la parte di varianza spiegata da questi modelli era più alta ( $R^2 = 0.4$  per Do1000 e Pe1000) rispetto all'analisi a livello di parola; tuttavia, ciò è dovuto soprattutto all'integrazione del fattore 'fonema' all'interno dei modelli; la parte di varianza spiegata dai valori di *SentiWordNet* ha raggiunto solo 0.004 e 0.007 per Do1000 e Pe1000 rispettivamente.

<b>Effetto</b>	<b>SentiWN&gt;0</b>	<b>SentiWN&lt;0</b>
<i>F0 min</i>	*	<i>ns</i>
<i>F0 max</i>	<i>ns</i>	***
<i>F0 range</i>	<i>ns</i>	***
<i>F0 mean</i>	*	***
<i>F0 stdev</i>	<i>ns</i>	***
<i>Shimmer</i>	<i>ns</i>	**
<i>Jitter</i>	<i>ns</i>	***
<i>HNR</i>	***	***
<i>Do1000</i>	***	***
<i>Pe1000</i>	***	***
<i>HAM</i>	<i>ns</i>	<i>ns</i>

**Tabella 4.** Effetto di *SentiWordNet\_pos* e *SentiWordNet\_neg* sui valori acustici.

## 4 Conclusioni

La correlazione tra indici lessicali e acustici del parlato letto emozionale sembra essere significativa, ma di portata ridotta, sia a livello di parola, sia a livello di frase. Gli indici di F0 sembrano essere influenzati dalla valenza della frase e della parola, ma la parte di varianza spiegata rimane ridotta. Tali risultati confermano ed estendono quanto riportato da Charfuelan & Schröder (2012) su dati di un solo audiolibro, in cui erano state osservate correlazioni moderate per indici di F0 ed energia.

I dati mostrano una grande quantità di variabilità inter-speaker: risulta evidente che i locutori utilizzano diversi indici acustici per esprimere stati emozionali. Inoltre, un limite della nostra analisi risiede nell'utilizzo (inevitabile, data la mole di dati analizzati) di trascrizioni e annotazione automatiche, i cui errori causano senza dubbio un certo tasso di rumore nei dati, riducendo le relazioni osservabili tra le diverse variabili studiate. Infine, l'assenza di punteggiatura nel corpus *LibriSpeech* rende impossibile (o molto complesso) differenziare tra discorso indiretto e diretto, nel quale ci si potrebbe aspettare un parlato più prettamente emozionale. Per il futuro, simili ipotesi potranno essere verificate su corpora più recenti costruiti con fini più specifici e adatti, come SynPaFlex (Sini et al., 2008).

Per concludere, riprendiamo il tema dell'interazione tra i vari livelli linguistici per l'espressione delle emozioni nel parlato. I risultati del nostro studio suggeriscono che i vari livelli linguistici analizzati (lessicale e acustico) sono relativamente slegati uno dall'altro per l'espressione delle emozioni. Questo significa

che, per una determinata frase, i locutori hanno tendenza ad affidare l'espressione dello stato emozionale a uno solo dei due livelli analizzati. Questo può essere vero soprattutto per il parlato letto, in cui il locutore non è coinvolto direttamente, soprattutto nel caso del narratore di un audiolibro. Dunque, l'utilizzo della SA per lo studio del parlato emozionale appare non del tutto appropriato per selezionare materiale emozionalmente marcato, in quanto si baserebbe sull'assunzione che gli indici lessicali e acustici di emozionalità vadano di pari passo e tendano a co-occorrere. Tuttavia, rimane da esplorare la correlazione tra variabili lessicali e acustiche per altri tipi di parlato, in particolar modo per il parlato spontaneo – in cui i locutori siano più direttamente coinvolti rispetto al contenuto semantico.

## Bibliografia

- Audibert N., Aubergé V., Rilliard A. 2005. The prosodic dimensions of emotion in speech: the relative weights of parameters. *Proc. of the Ninth European Conference on Speech Communication and Technology*, 4-8 September 2005, Lisbon, Portugal.
- Baccianella S., Esuli A., Sebastiani F. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. of LREC*, 17-23 May, Valletta, Malta, pp. 2200-2204.
- Banse R., Scherer K.R. 1996. Acoustic profiles in vocal emotion expression, *Journal of personality and social psychology*, vol. 70, no. 3, 614–636.
- Bates D., Maechler M., Bolker B., Walker S. 2014. Fitting Linear Mixed-Effects Models Using lme4, *Journal of Statistical Software*, vol. 67, no. 1, 1-48.
- Boersma P., Weenink D.. 2018. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.37, retrieved 3 February 2018 from <http://www.praat.org/>
- Burkhardt F., Sendlmeier W.F. 2000. Verification of acoustical correlates of emotional speech using formant-synthesis. In *SpeechEmotion-2000*, pp. 151-156.
- Charfuelan M., Schröder M. 2012. Correlation analysis of sentiment analysis scores and acoustic features in audiobook narratives. In *Proc. of the 4th International Workshop on*

- Corpora for Research on Emotion Sentiment & Social Signals (ES3)*, 26 May 2012, Istanbul, Turkey, pp. 99-103.
- Drioli C., Tisato G., Cosi P., Tesser F. 2003. Emotions and voice quality: experiments with sinusoidal modeling. In *Proc. of the Voice Quality: Functions Analysis and Synthesis (VOQUAL) Workshop*, 27-29 August, Geneva, Switzerland.
- Ekman P. 2000. Basic Emotions. In T. Dalgleish and T. Power (eds.) *Handbook of Cognition and Emotion*, 39.6, London (UK), John Wiley & Sons, pp. 45-60.
- Gilbert C.J., Hutto E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proc. of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, 2-4 June, Ann Arbor MI, US.
- Hammarberg B., Fritzell B., Gauffin J., Sundberg J., Wedin L. 1980. Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, vol. 90, 441-451.
- Johnstone T., Scherer K.R. 2000. Vocal communication of emotion. In M. Lewis and J. Haviland (eds.) *Handbook of emotions 2*, London-New York: The Guildford Press, pp. 220-235.
- Liu B. 2012. Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no.1, 1-167, 2012.
- McGlohon M., Glance N., Reiter Z. 2010. Star quality: Aggregating reviews to rank products and merchants. In *Proc. of the International Conference on Weblogs and Social Media (ICWSM-2010)*, 23-26 May, Washington DC, US.
- Mohammad S. 2011. From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. In *Proc. of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, 24 June, Oregon, US.
- Mohammad S., Yang T. 2011. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. In *Proc. of the ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA-2011)*, 24 June, Alicante, Spain.
- Panayotov V., Chen G., Povey D., Khudanpur S. 2015. Librispeech: an ASR corpus based on public domain audio books. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 19-24 April, Brisbane, Australia.
- Reilly J., Seibert L. 2003. Language and emotion. In R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.), *Handbook of affective sciences*, OUP, pp. 535-559.
- Russell J.A. 1980. A circumplex model of affect, *Journal of personality and social psychology*, vol. 39, no.6, 1161-1178.
- Sadikov E., Parameswaran A., Venetis P. 2009. Blogs as predictors of movie success. In *Proc of the Third International Conference on Weblogs and Social Media (ICWSM-2009)*, 17-20 May, San Jose, CA, US.
- Scherer K.R.. 1989. Vocal correlates of emotion. In A. Manstead and H. Wagner (eds.) *Handbook of psychophysiology: Emotion and social behavior*, London: Wiley, pp. 165-197.
- Scherer K.R. 2002. Vocal communication of emotion: A review of research paradigms, *Speech Communication*, vol. 40, 227-256.
- Schröder M., Cowie R., Douglas-Cowie E., Westerdijk M., Gielen S. 2001. Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proc. of EUROSPEECH 2001 – Seventh European Conference on Speech Communication and Technology*, 3-7 September 2001, Aalborg, Denmark.
- Sini A., Lolive D., Vidal G., Tahon M., E. Delais-Roussarie. 2008. SynPaFlex-Corpus: An Expressive French Audiobooks Corpus Dedicated to Expressive Speech Synthesis. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 7-12 May 2018, Miyazaki (Japan), pp. 4289-4296.
- Tsiakoulis P., Raptis S., Karabetsos S., Chalamandaris A. 2016. Affective word ratings for concatenative text-to-speech synthesis. In *Proc. of the 20th Pan-Hellenic Conference on Informatics*, 10-12 November, Patras, Greece.