

GLEND A GURRADO

Sulla codifica e decodifica della sorpresa

The production and perception of surprise

The study investigates the perception and production of surprise. To date very few works have analysed the vocal communication of surprise compared to other fundamental emotions. Our research aims to verify the phonetic indices of surprise, distinguishing between positive and negative attitude. Furthermore we aim to analyse the role played by some parameters in the perception of the level of surprise.

The findings obtained by the acoustical analysis suggested that surprise presents higher values than neutral communication in terms of F0, intensity and duration and that positive surprise is more prosodically marked than negative surprise. Finally, the perception test showed that F0 and duration manipulations seem not to significantly affect the decoding of surprise.

Keywords: communication of emotions, surprise, prosody, acoustic analysis, perception.

1. *Introduzione*

Il concetto di senso di sorpresa è solitamente associato all'idea di cogliere qualcuno impreparato attraverso un'azione inaspettata, che desta stupore. La sorpresa è annoverata fra le emozioni base tra cui rientrano anche gioia, tristezza, paura, disgusto e rabbia. Le emozioni fondamentali sono selezionate in base ai dati relativi al riconoscimento universale delle emozioni per mezzo delle espressioni facciali e vocali (tra gli altri, Tomkins, 1962, 1963; Izard, 1994; Ekman, 2003). Nella maggior parte dei contributi dedicati ai correlati acustici delle emozioni, la sorpresa è meno indagata rispetto alle altre e, spesso, addirittura esclusa (fra gli altri, Juslin, Laukka, 2003; Scherer, 2003). Secondo alcuni studi, questa emozione, assieme al disgusto, è difficilmente identificata tramite la voce (Scherer, 1989): evidentemente queste emozioni prediligono un tipo di comunicazione diversa da quella vocale, un comportamento, secondo gli autori, consolidatosi nel corso dell'evoluzione (Johnstone, Scherer, 2000). Tuttavia, contrariamente a questa teoria, in alcuni lavori sperimentali, l'emozione della sorpresa è in buona percentuale riconosciuta dagli uditori (Cahn, 1990; Abelin, Allwood, 2000; Iriando, Guas, Rodríguez, Lázaro, Montoya, Blanco, Bernadas, Oliver, Tena & Longhi, 2000). In alcuni contesti la sorpresa è stata, però, confusa con altre emozioni, come la gioia e la rabbia, con cui condivide una serie di caratteri prosodici (Cahn, 1990). La maggior parte degli studi concorda nell'associare la sorpresa ad alti livelli frequenziali, una certa variabilità di F0 e un'intensità maggiore rispetto al parlato non marcato. Relativamente all'italiano, la ricerca sulla comunicazione delle emozioni è in crescita, tuttavia i contributi che considerano

l'emozione della sorpresa sono ancora pochi (es., De Marco, Paone, 2014; Maffia, Pellegrino & Pettorino, 2014; Gili Fivela, 2015).

Alla luce di quanto rilevato da tali studi, emerge una riflessione: la sorpresa, rispondendo a stimoli contestuali o ambientali di natura diversa, può avere una connotazione positiva o negativa, a seconda che l'evento sia considerato vantaggioso o dannoso per il parlante; questa emozione può essere associata a caratteristiche psicologiche distinte ed espressa tramite mezzi diversi in relazione all'evento scatenante e al modo in cui il soggetto lo valuta. Tuttavia, finora, pochissimi dei lavori incentrati (non specificamente) sulla realizzazione acustica del senso di sorpresa si sono interrogati sulla distinzione fra le diverse connotazioni, positiva e negativa, di questa emozione. La presente ricerca intende, dunque, indagare la produzione e la percezione della sorpresa, con l'intento di verificare in primis le differenze prosodiche fra sorpresa e parlato neutrale e in secondo luogo quelle fra sorpresa positiva e sorpresa negativa. In aggiunta, una verifica percettiva ha lo scopo di comprendere il ruolo svolto da alcuni parametri acustici nella decodifica dei diversi livelli di sorpresa espressi dal parlante.

2. Un'emozione base

La letteratura dedicata alla comunicazione delle emozioni fa riferimento a due tradizioni teoriche principali: la *Discrete Theory* e l'approccio dimensionale. La prima prevede l'esistenza di un certo numero di emozioni fondamentali caratterizzate da specifiche risposte fisiologiche ed espressioni facciali e vocali (tra gli altri, Tomkins, 1962, 1963; Izard, 1994; Ekman, 2003). Ekman (2003) ha messo in luce che soltanto sei emozioni sono riconosciute a livello universale, si parla oggi, infatti, delle *big six* ossia sorpresa, gioia, tristezza, paura, disgusto e rabbia; tuttavia, a seconda dell'approccio adottato, il numero delle emozioni ritenute fondamentali può variare. Come suggerito da Scherer (2003), però, i sostenitori della *Discrete Emotion Theory* non hanno fornito predizioni concrete relative ai *pattern* vocali delle emozioni base. Le emozioni possono essere comunicate sia attraverso mezzi espressivi corporali (espressioni facciali, gestualità, movimenti del corpo) che vocali: alcune emozioni propendono per un tipo di comunicazione rispetto all'altro a causa del diverso modo di reagire agli stimoli esterni del nostro organismo; tali reazioni si consolidano a livello culturale e sociale, portando a una cristallizzazione delle associazioni fra emozioni e risposte fisiologiche:

Concretely, the assumption is that the emotional arousal of the speaker is accompanied by physiological changes that will affect respiration, phonation, and articulation in such a way as to produce emotion-specific patterns of acoustic parameters [...] (Scherer, 2003: 229).

Secondo l'approccio dimensionale, per il quale si fa riferimento principalmente a Wundt (1874/1905), le emozioni che stimolano un certo tipo di reazioni a un certo grado sono accumulate dall'attivazione dei medesimi *pattern* prosodici. Una

delle dimensioni maggiori a cui fa riferimento questo approccio si basa su livelli contrapposti di attivazione/eccitamento: a seconda del grado di attivazione delle alterazioni fisiologiche che stimolano, si distingue fra emozioni ad alta attivazione ed emozioni a bassa attivazione. Le prime (paura, gioia, sorpresa e rabbia) sono generalmente caratterizzate da alti valori di F0, un'estensione tonale ampia, un'intensità alta, una maggiore velocità di eloquio e quindi una riduzione del numero di pause; diversamente, le seconde (tristezza e disgusto) sono veicolate da bassi valori di F0, un *range* tonale compresso, intensità bassa, un eloquio più lento e lunghe pause. Una seconda dimensione è quella relativa alla valenza in cui si contrappongono emozioni piacevoli-spiacevoli. Infine, la dimensione della potenza riguarda l'idea di sottomissione o dominanza derivante dal raggiungimento di uno scopo¹.

3. *Difficoltà di riconoscimento della sorpresa*

Uno degli argomenti ancora non adeguatamente esplorato riguarda le diverse modalità in cui le emozioni sono riconosciute; in altre parole, alcune emozioni sono più facilmente identificate tramite le espressioni facciali rispetto ad altre che prediligono la voce. Tristezza e rabbia sono, ad esempio, distinte correttamente tramite espressioni vocali rispetto al disgusto; questa emozione non presenta, infatti, percentuali di riconoscimento vocale convincenti. Probabilmente, nel corso dell'evoluzione l'espressione del disgusto ha stimolato un tipo di espressione per lo più connessa alla mimica e meno al coinvolgimento delle risorse vocali (Johnstone, Scherer, 2000): di conseguenza è più difficile riconoscere questa emozione affidandosi esclusivamente ai pattern acustici a essa accostati.

[...] the voice is probably more suited to the expressive and communicative needs of certain emotions than of others. For instance, there is a clear adaptive advantage to being able to warn (fear) or threaten (anger) others in a fairly indirect way over large distances – something for which vocal expression is ideally suited. [...] In contrast, naturally occurring vocal expressions of disgust probably consist of brief affect bursts or vocal emblems (e.g., “Yuck!”) rather than long sentences spoken with a “disgust-specific” voice quality (Johnstone, Scherer, 2000: 229).

Tale assunto fa propendere più per la tesi darwiniana secondo cui, anche se è possibile individuare delle tendenze di espressione tramite le quali raggruppare le emozioni, tuttavia la comunicazione delle emozioni è da considerarsi un fenomeno variabile (Darwin, 1872). L'ipotesi avanzata da Johnstone, Scherer (2000) è estesa anche alla sorpresa: la corretta codifica e decodifica di questa emozione sembrerebbe essere affidata per lo più alla mimica del volto. Tuttavia, le deduzioni avanzate dagli studiosi non sono supportate da dati sufficienti a sostenere una conclusione di questo tipo: i dati riportati da Scherer (2003: 236) mostrano che la maggior parte degli studi che annovera la sorpresa fra le emozioni indagate ha analizzato l'identificazione di questa emozione esclusivamente tramite le espressioni facciali.

¹ Per una rassegna degli studi dedicati all'argomento cfr. Scherer (2003).

Al fine di chiarire meglio la questione relativa alle modalità di riconoscimento della sorpresa, nel paragrafo che segue si cercherà in primis di delineare un profilo prosodico di questa emozione e, inoltre, di verificare le percentuali di identificazione di questa da parte degli uditori.

4. *I correlati acustici*

Nella maggior parte degli studi dedicati ai correlati acustici delle emozioni² la sorpresa è meno indagata rispetto alle altre e, in alcuni casi, addirittura esclusa (fra gli altri, Juslin, Laukka, 2003; Scherer 2003). Tuttavia, in alcuni lavori la sorpresa compare fra le emozioni analizzate a livello di produzione e percezione. In questo paragrafo saranno passati in rassegna alcuni contributi specificamente dedicati alla codifica e alla decodifica delle emozioni, in cui è considerata anche la sorpresa. Relativamente all'inglese – con l'intento primario di testare la possibilità di riprodurre l'espressione delle emozioni tramite sintesi vocale – un test percettivo effettuato in Cahn (1990) ha messo in luce che la sorpresa, così come le altre emozioni base considerate, è stata riconosciuta nel 50% dei casi; in alcuni contesti, la sorpresa è stata, tuttavia, confusa con altre emozioni, come la gioia e la rabbia, con cui condivide alcuni caratteri prosodici. Con riferimento allo svedese, in Abelin, Allwood (2000) – uno studio specificamente dedicato all'osservazione della corrispondenza fra emozioni e prosodia – è dimostrato che la sorpresa è caratterizzata da una forte variabilità di F0, un'alta intensità e una maggior durata rispetto alle altre emozioni base considerate. Il contributo ha, inoltre, messo in luce che la sorpresa è stata identificata correttamente dai partecipanti all'indagine sperimentale nella maggior parte dei casi. In Iriondo et al. (2000) sono indagate le modalità di espressione vocale delle emozioni in spagnolo, la validazione del modello usato è realizzata mediante sintesi vocale. L'analisi acustica ha dimostrato che l'emozione della sorpresa è identificata da più della metà degli uditori; questa emozione stimola l'attivazione di alcuni parametri prosodici rispetto al parlato neutrale, mostrando un aumento della F0, un incremento della variabilità di F0, un'intensità maggiore e una diminuzione della durata dei gruppi fonici. Nel contributo di Belin, Fillion-Bilodeau & Gosselin (2008), dedicato alla produzione e alla percezione dei *nonverbal affective bursts*, con riferimento a dei dati raccolti da francofoni, la sorpresa è caratterizzata da valori medi, minimi e massimi di F0 maggiori rispetto al parlato neutrale, così come la deviazione standard. In particolare, la media dei valori di F0x registrati per la sorpresa, assieme a quella della paura, supera significativamente i valori raggiunti dalle altre emozioni. L'intensità media è, invece, comparabile a quella del parlato neutrale, mentre la deviazione standard è più alta. La sorpresa è stata riconosciuta nel 75% dei casi, tuttavia, è stata più volte confusa con la paura.

Pell, Paulmann, Dara, Alasseri & Kotz (2019) analizzano la comunicazione delle emozioni in produzione e in percezione, con riferimento a quattro lingue (inglese,

² Per un'analisi puntuale dei correlati acustici delle emozioni base cfr., fra gli altri, Juslin, Laukka (2003).

tedesco, hindi e arabo). Focalizzandosi sull'analisi di una serie di *pseudo-sentences* prodotte da un gruppo di parlanti madrelingua con una minima esperienza attoriale, gli autori hanno rilevato che la comunicazione della sorpresa è caratterizzata da una $F0x$ maggiore e un *range* frequenziale più ampio non solo rispetto al parlato neutrale ma anche alle altre emozioni base in tutte e quattro le lingue indagate. La sorpresa viene, inoltre, espressa con un eloquio mediamente veloce, con alcune differenze a seconda della lingua di riferimento.

Relativamente all'italiano, la ricerca sulla comunicazione delle emozioni è in crescita, tuttavia gli studi che considerano l'emozione della sorpresa sono ancora pochi. Alcune informazioni si ricavano dai contributi incentrati sulla produzione e sulla percezione delle emozioni in italiano L2. In Maffia et al. (2014), l'analisi acustica di una serie di stimoli raccolti tramite *CardTask* ha confermato che la sorpresa innesca un innalzamento dei valori di $F0$ e un'estensione del *range* tonale. Lo studio di De Marco, Paone (2014), che è incentrato sulla codifica e la decodifica delle sei emozioni primarie in apprendenti italiano L2, ha confermato che la sorpresa presenta le stesse caratteristiche prosodiche delle altre emozioni ad alta attivazione; per quanto concerne la percezione, è emerso che la sorpresa è identificata correttamente nella maggior parte dei casi (si rilevano, però, difficoltà più o meno significative a seconda della lingua nativa degli uditori). Nel lavoro di Gili Fivela (2015) è indagato il ruolo delle informazioni visive rispetto a quello delle caratteristiche prosodiche nell'interpretazione del parlato, in particolare, con riferimento alla possibilità di una diversa integrazione di tali informazioni in base alla marcatezza degli enunciati. L'autrice analizza, oltre a domande e assertive, anche le esclamative che esprimono sorpresa con valenza positiva e negativa. L'analisi prosodica degli stimoli audio, raccolti nel test di produzione mediante *Discourse Completion Task*, ha messo in luce che le esclamative veicolanti l'emozione sorpresa sono caratterizzate da valori di intensità più alti rispetto alle altre tipologie frasali non marcate emotivamente; infine, i valori dell'intervallo di frequenze sono particolarmente variabili, una tendenza probabilmente ascrivibile alle strategie impiegate da ciascun parlante.

5. *La ricerca*

La sorpresa appare, dunque, un'emozione ibrida, dai confini sfumati che la conducono a confondersi in alcuni contesti con la gioia, in altri con la rabbia. La sorpresa, rispondendo a stimoli di natura diversa, può avere una connotazione positiva o negativa, a seconda che l'evento sia considerato vantaggioso o dannoso, piacevole o spiacevole, per il parlante: questa emozione può essere associata a caratteristiche psicologiche distinte ed espressa tramite mezzi diversi in relazione all'evento scatenante e al modo in cui il soggetto lo valuta. Tuttavia, finora, gli studi incentrati sulla realizzazione acustica del senso di sorpresa hanno dedicato poca attenzione alle diverse connotazioni, positiva e negativa, di questa emozione. La presente ricerca intende, dunque, indagare la produzione e la percezione della sorpresa, tenendo conto anche della sua natura dicotomica. Specificamente ci si chiede:

1. Quali sono i correlati acustici della sorpresa?
2. La sorpresa positiva differisce da quella negativa in termini prosodici?
3. Da quali parametri prosodici è influenzata la percezione della sorpresa?

La novità della ricerca consiste in primis nell'osservazione degli indici fonetici che in produzione differenziano la sorpresa positiva da quella negativa e in secondo luogo nello svolgimento di una verifica percettiva volta non alla valutazione del riconoscimento dell'emozione convogliata ma all'identificazione degli aspetti prosodici più coinvolti nella decodifica del grado di sorpresa. Al fine di rispondere a tali quesiti, la ricerca è stata organizzata in due fasi: la prima (codifica) verte sull'analisi acustica di un corpus di frasi che esprimono sorpresa, positiva e negativa, pronunciate da un gruppo di parlanti baresi; la seconda (decodifica) è incentrata all'analisi percettiva di un sottocampione di frasi target estratte dal primo esperimento.

5.1 I partecipanti

Un gruppo di dieci soggetti residenti a Bari è stato scelto per il nostro esperimento. I parlanti, cinque uomini e cinque donne fra i 25 e i 35 anni, sono studenti universitari o laureati di pari provenienza ed estrazione socioculturale. Tutti hanno confermato di essere nati e aver vissuto prevalentemente a Bari. Nessuno dei partecipanti possiede nozioni di fonetica o fonologia.

5.2 I materiali

I materiali sono composti da 20 sceneggiature ciascuna formata da un contesto in situazione, che dovrebbe suscitare sorpresa nel parlante, e una frase target. Di queste, dieci esprimono sorpresa positiva, poiché inserite in un contesto in cui è descritto un evento gradito, e dieci comunicano sorpresa negativa, poiché precedute dalla descrizione di un fatto spiacevole. La tipologia frasale scelta è l'esclamativa, che per definizione esprime la sorpresa del parlante rispetto alla realizzazione di un evento inaspettato: il significato della modalità esclamativa risiede proprio nella comunicazione di questo approccio del locutore alla realtà dei fatti. Nello specifico, si tratta dell'esclamativa non *wh*: con tale dicitura in questa sede intendiamo principalmente quelle strutture assertive che convogliano un significato esclamatorio, es. *Martina ha trovato un lavoro!*. Nel nostro campione sono presenti costruzioni formate dalla successione SNsogg+SV (es. *Gianni ha quasi settant'anni!*) e SV+SNsogg (es. *Ha telefonato il preside della tua scuola!*). Come controllo, è stato predisposto un campione di 20 frasi non connotate emotivamente, di tipologia frasale assertiva, sintatticamente identiche alle frasi esclamative target.

5.3 La sessione di registrazione

Ogni partecipante è stato coinvolto singolarmente in due sessioni da 30 minuti, intervallate da una pausa della stessa durata. Durante l'incontro ogni speaker è stato invitato a leggere le sceneggiature proposte a gruppi di cinque, dapprima in maniera silenziosa – al fine di evitare una lettura artificiosa – e poi ad alta voce, nel modo più

naturale possibile: è stato suggerito al lettore di immedesimarsi nel parlante coinvolto nella situazione comunicativa descritta dal contesto di ciascuna sceneggiatura. Per non influenzare la lettura delle frasi target, non è stata fornita alcuna informazione relativa alla finalità della ricerca. L'audioregistrazione dei materiali è avvenuta in un ambiente insonorizzato tramite un Tascam DR-40 (acquisizione formato .WAV, con frequenza di campionamento pari a 44 KHz e risoluzione a 24 bit). Con lo scopo di assicurare un livello di intensità corretto e costante, ciascun partecipante è stato invitato a posizionarsi a una distanza di circa 20 cm dal microfono. Il test ha prodotto 400 enunciati bersaglio, di cui 200 esclamative non *wh* (100 sorpresa positiva + 100 sorpresa negativa) e 200 assertive.

6. Produzione

Il disegno sperimentale dello studio ha previsto due fasi consistenti la prima in un'analisi acustica, la seconda in un'analisi percettiva. L'analisi acustica, volta a individuare i tratti che caratterizzano la sorpresa rispetto al parlato non emotivo, ha inteso inoltre verificare se la sorpresa positiva differisse da quella negativa in termini prosodici.

6.1 Il protocollo

Per la verifica acustica, sono stati considerati i seguenti parametri: escursione melodica convertita in semitoni (ST), valore frequenziale dell'Onset e dell'Offset dell'enunciato (Hz)³, valore dell'intensità media (dB), durata totale dell'intero enunciato (ms), durata dell'ultima vocale tonica⁴ (ms), velocità di eloquio (sill/s). L'analisi è stata effettuata tramite il software *PRAAT* (Boersma, Weenink, 2018). La significatività statistica, impostata per $p < 0,05$, è stata svolta con l'impiego del *Paired Sample T test*.

6.2 I risultati

I dati raccolti hanno rivelato che gli enunciati che esprimono sorpresa presentano un'estensione tonale, un'intensità media e una durata maggiore rispetto agli enunciati neutrali. In aggiunta, la sorpresa positiva è più marcata a livello prosodico rispetto alla sorpresa negativa.

I risultati saranno presentati nei paragrafi successivi separatamente per parametro indagato.

³ Per estrarre i valori di Onset e Offset sono stati misurati rispettivamente i livelli frequenziali medi dei primi e degli ultimi 30 ms del contorno intonativo.

⁴ Si è deciso di misurare la durata dell'ultima vocale tonica poiché è l'elemento strutturale su cui si realizza il nucleo tonale: si tratta, quindi, della vocale più forte dal punto di vista metrico, secondo quanto postulato dai principi del modello autosegmentale e metrico. Inoltre, in prospettiva pragmatica, è anche l'elemento su cui si realizza l'illocuzione e quindi si compie la forza comunicativa della frase.

6.2.1 Frequenza Fondamentale

Dai nostri risultati è emerso che la sorpresa è caratterizzata da un'estensione tonale più ampia rispetto a quanto riscontrato nel campione di controllo: la differenza media è di 2 ST ($t = 5,773$; $df = 199$; $p < 0,001$). In questa sede sono stati considerati anche i valori frequenziali dei primi e degli ultimi 30 ms dell'enunciato. Dai dati è emerso che nelle esclamative non *wh* che esprimono sorpresa l'Onset raggiunge un livello frequenziale di 3 ST ($p < 0,001$) più alto rispetto alla controparte non connotata emotivamente; mentre la differenza relativa ai valori dell'Offset è di 1,2 ST ($p < 0,001$). Nelle Figg. 1 e 2 sono riportati gli spettrogrammi relativi alla frase "Martina ha trovato un lavoro" pronunciate con senso di sorpresa (Fig. 1) e in modo non marcato emotivamente (Fig. 2) dal parlante GP. Come si può notare, la frase che esprime sorpresa presenta un contorno molto più modulato: in particolare, un Onset più alto (+18,7 ST) e un *pitch range* più esteso (+12,2 ST).

Relativamente al confronto tra sorpresa positiva e sorpresa negativa i risultati rivelano che quest'ultima è caratterizzata da un *range* tonale più compresso rispetto alla prima: specificamente si tratta di una differenza di 2,9 ST ($t = 5,756$; $df = 99$; $p < 0,001$). Per quanto concerne l'Onset di F0, la sorpresa positiva rivela valori significativamente più alti rispetto alla sorpresa negativa (+1,3 ST; $p < 0,001$). Diversamente, il confronto relativo ai valori di Offset non è significativo (+0,2 ST; $p > 0,05$). Gli spettrogrammi riportati nelle Figg. 3 e 4 rappresentano un esempio di queste tendenze. La frase "Luca è arrivato" (Fig. 3), che esprime sorpresa positiva, presenta dei valori frequenziali più alti rispetto alla frase "Mara è partita!" che comunica il senso di sorpresa negativo provato dal parlante. Il primo contorno è più modulato e presenta una prominenza sul verbo.

Figura 1 - *Forma d'onda, spettrogramma a banda larga e curva di F0 della frase che esprime sorpresa "Martina ha trovato un lavoro!" pronunciata dal locutore GP.*
Onset: 280,2 Hz, PR: 21,2 ST

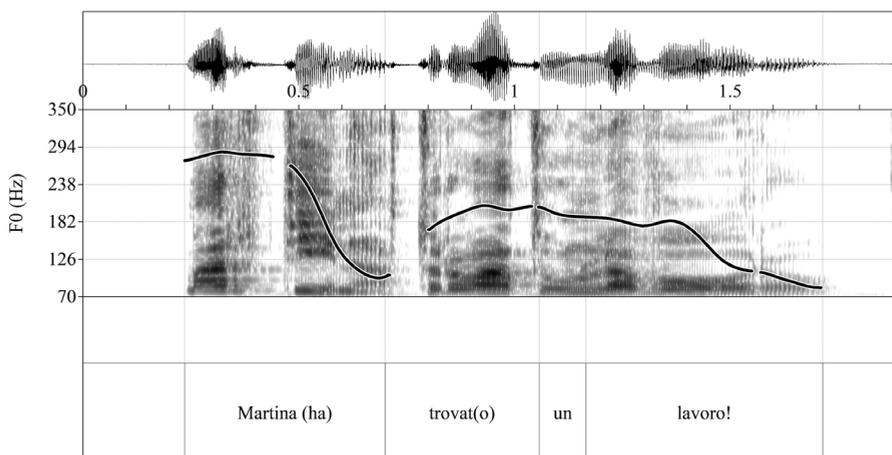


Figura 2 - *Forma d'onda, spettrogramma a banda larga e curva di F0 della frase non connotata emotivamente "Martina ha trovato un lavoro." pronunciata dal locutore GP. Onset: 94,7 Hz, PR: 9 ST*

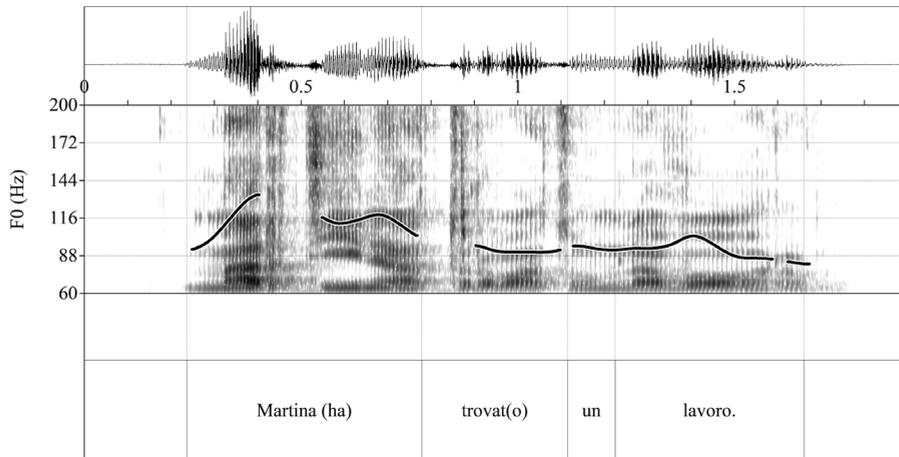


Figura 3 - *Forma d'onda, spettrogramma a banda larga e curva di F0 della frase che esprime sorpresa positiva "Luca è arrivato!" pronunciata dal locutore GP. Onset: 245,4 Hz, PR: 21,6 ST*

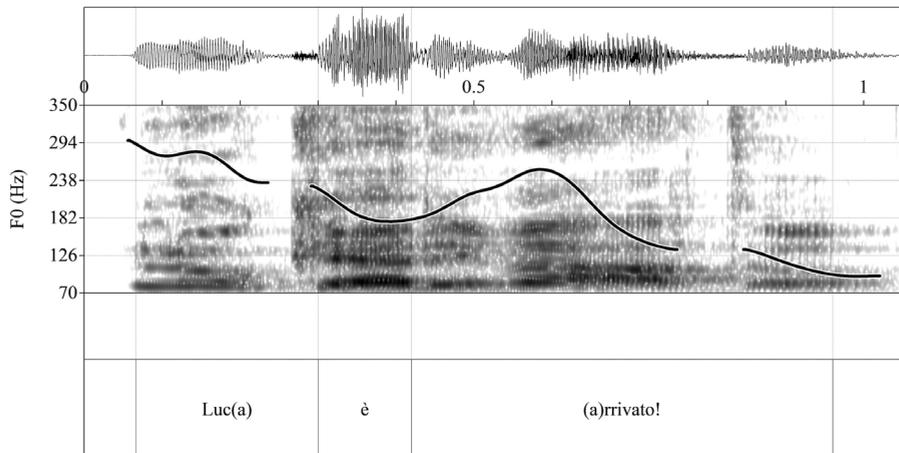
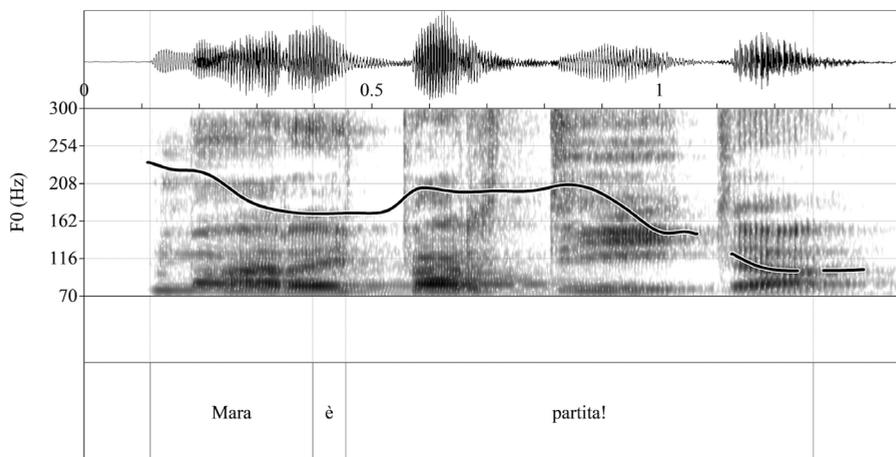


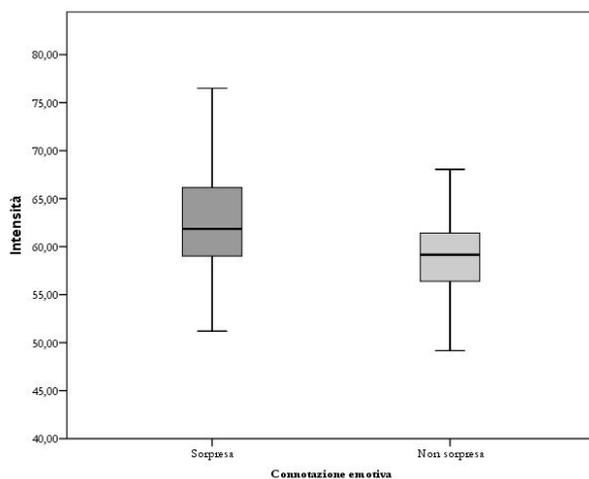
Figura 4 - Forma d'onda, spettrogramma a banda larga e curva di F0 della frase che esprime sorpresa negativa "Mara è partita!" pronunciata dal locutore GP. Onset: 157,2 Hz, PR: 11,3 ST



6.2.2 Intensità

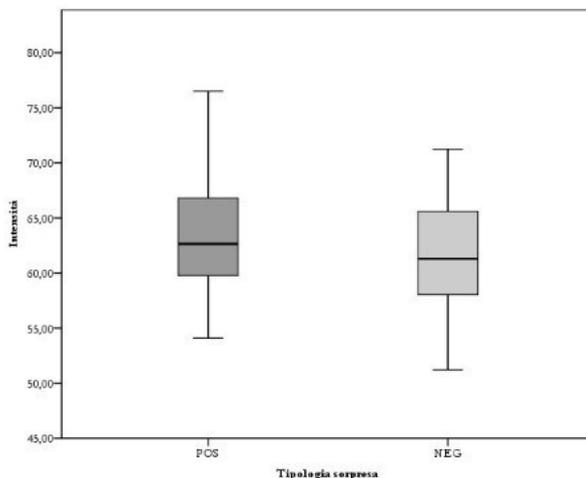
Gli enunciati che comunicano sorpresa sono anche più intensi rispetto agli enunciati non emotivi. Come si può notare dal *boxplot* in Fig. 5, si rileva una differenza significativa di 3,9 dB ($t = 12,991$; $df = 199$; $p < 0,01$).

Figura 5 - Intensità media (dB) delle frasi che esprimono sorpresa e di quelle non connotate emotivamente



La sorpresa positiva rivela un'intensità media maggiore rispetto alla sorpresa negativa (+2 dB, $t = 3,755$; $df = 99$; $p < 0,001$), come riportato nel *boxplot* in Fig. 6.

Figura 6 - Intensità media (dB) delle frasi che esprimono sorpresa positiva e di quelle che comunicano sorpresa negativa



6.2.3 Durata

Relativamente a questo parametro abbiamo registrato sia la durata totale dell'intero enunciato che la durata dell'ultima vocale tonica. La sorpresa rivela una durata totale maggiore rispetto al campione di controllo: la differenza fra i due gruppi è di 174,2 ms ($t = 12,179$; $df = 199$; $p > 0,001$). La vocale nucleare è invece più lunga di 34,2 ms ($t = 11,540$; $df = 199$; $p < 0,001$). Gli enunciati che esprimono sorpresa sono quindi più lenti (-0,8 ms, $t = -13,207$; $df = 199$; $p < 0,001$). Nelle Figg. 7 e 8, è possibile notare come la parlante GM pronunci la frase che esprime sorpresa *Luca è arrivato!* (Fig. 7) allungando la vocale nucleare di circa 80 ms rispetto alla sua controparte non connotata emotivamente (Fig. 8).

Relativamente al confronto fra sorpresa positiva e sorpresa negativa, è necessario specificare che le frasi appartenenti ai due gruppi messi a confronto non sono morfologicamente e sintatticamente identiche e ciò potrebbe aver inficiato la valutazione delle tendenze temporali. Ciò nonostante, dato che in fase di composizione delle sceneggiature sono state scelte frasi target formate da un numero di costituenti simile, è possibile comunque analizzare e confrontare i risultati. I dati hanno mostrato che il parametro della durata non è influenzato dall'attitudine positiva o negativa; in questo caso le differenze fra i due gruppi non sono significative (-55 ms, $t = -1,118$; $df = 99$; $p > 0,05$). Relativamente alla durata dell'ultima vocale tonica, si registra una differenza di soli 4,1 ms ($t = 0,644$; $df = 99$; $p > 0,05$). Tuttavia, il confronto relativo alla velocità elocutiva è significativo (-0,36, $t = -2,599$; $df = 99$; $p < 0,05$).

Figura 7 - Forma d'onda, spettrogramma a banda larga e curva di F0 della frase che esprime sorpresa "Luca è arrivato!" pronunciata dalla locutrice GM.
DT: 1160 ms, DVT: 224 ms

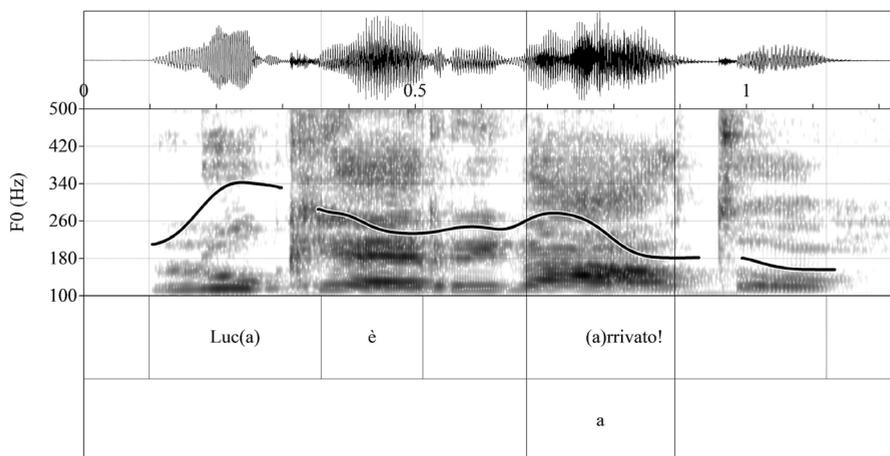
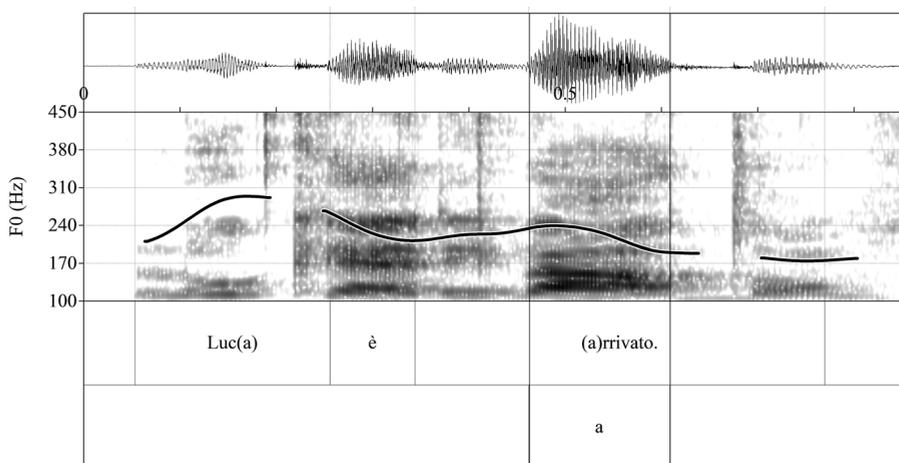


Figura 8 - Forma d'onda, spettrogramma a banda larga e curva di F0 della frase non connotata emotivamente "Luca è arrivato!" pronunciata dalla locutrice GM.
DT: 1030 ms, DVT: 146 ms



7. Percezione

Dai dati raccolti nell'analisi acustica, è emerso che le frasi che esprimono sorpresa presentano un'estensione tonale più ampia, un Onset di F0 più alto, un'intensità maggiore e un'eloquio più lento. Con l'obiettivo di verificare il ruolo svolto da alcuni di questi parametri nella percezione della sorpresa, è stato elaborato un test percettivo.

7.1 Il protocollo

Il principale obiettivo è quello di comprendere la misura in cui F0 e durata influenzino il giudizio degli ascoltatori nella percezione del grado di sorpresa provato dal parlante; in particolare, i parametri indagati sono l'Onset di F0 e la durata dell'ultima vocale tonica (DVT). In questo caso, si è ipotizzato che all'abbassamento dei valori di questi parametri corrisponda una riduzione del livello di sorpresa percepito. L'esperimento ha previsto la manipolazione di un sottocorpus di stimoli preselezionati: due esclamative originali pronunciate dalla parlante CL sono state sottoposte all'abbassamento in tre step dei valori di Onset di F0 e della DVT. Gli enunciati scelti esprimono sorpresa positiva (CL_1T: *Luca è arrivato!*) e sorpresa negativa (CL_3T: *Mara è partita!*)⁵. La manipolazione dei parametri è stata effettuata tramite sintesi vocale delle frasi, mediante il sistema PSOLA disponibile nel software PRAAT. Specificamente, il procedimento sperimentale ha comportato: 1) l'accorciamento temporale della vocale nucleare in tre step da 15 ms; 2) l'abbassamento frequenziale dei primi 30 ms del contorno (Onset) in tre step da 1,5 ST⁶. La manipolazione è stata effettuata sia singolarmente che in maniera congiunta. Come si può dedurre dalla Tab. 1, gli stimoli ottenuti sono di sei tipi:

Tabella 1 - *Step di manipolazione singola e congiunta degli stimoli audio selezionati*

D1	D2	D3
-15 ms	-30 ms	-45 ms
F01	F02	F03
-1,5 ST	-3 ST	-4,5 ST
D1+F01	D2+F02	D3+F03
-15 ms -1,5 ST	-30 ms -3 ST	-45 ms -4,5 ST

Il numero totale degli stimoli ottenuti è 36, organizzati nel seguente modo:

- 6 esclamative con manipolazione della DVT;
- 6 esclamative con manipolazione dell'Onset;
- 6 esclamative con manipolazione congiunta di DVT e Onset;
- 18 *filler*⁷.

⁵ Questi enunciati sono stati selezionati in seguito a un pre-test in cui è stato chiesto a un gruppo ristretto di ascoltatori di giudicare, tramite scala Likert a 7 punti, il grado di sorpresa comunicato da una serie di frasi. Le esclamative CL_1T e CL_3T comparivano fra quelle che avevano raggiunto i punteggi più alti. La frase CL_1T (*Luca è arrivato!*) è accompagnata da un contesto che descrive l'arrivo ad una festa di Luca, un caro amico del parlante, si tratta quindi di un evento gradito. La situazione in cui è pronunciata CL_3T (*Mara è partita!*) è, invece, sgradita, poiché l'amica Mara è partita improvvisamente senza avvisare.

⁶ I valori acustici degli stimoli hanno dettato la scelta degli step: la media della differenza fra l'Onset e la DVT delle frasi che esprimono sorpresa da un lato e l'Onset e la DVT delle frasi non marcate emotivamente dall'altro è rispettivamente di circa 4 ST e 40 ms.

⁷ I *filler* consistono in frasi appartenenti a modalità frasali diverse da quella esclamativa, nello specifico assertive e interrogative non connotate emotivamente.

Gli stimoli ottenuti dalla sintesi, unitamente agli originali e ai *filler*, sono stati proposti in via telematica e in ordine randomizzato a un gruppo di 35 studentesse dell'Università degli Studi di Bari (età media 23,5 anni), nate e cresciute in area Apulo-Barese. Per ciascun enunciato si chiedeva "Secondo te la frase ascoltata esprime sorpresa?"; le partecipanti sono state quindi invitate a valutare il grado di sorpresa espresso tramite una scala Likert a 5 punti da "Per nulla" (1) a "Moltissimo" (5). Agli informatori non sono state fornite indicazioni circa le finalità della ricerca, allo scopo di non influenzare le loro risposte nel test.

7.2 I risultati

Da un'analisi preliminare dei punteggi raccolti, fra le due frasi selezionate è emersa una certa differenza di giudizi del livello di sorpresa percepito. Come riportato nella Tab. 2, i punteggi differiscono significativamente fra i due gruppi per tutti gli stimoli manipolati sia in maniera singola che congiunta. Come si può notare dai risultati, la media dei giudizi relativi alla frase che esprime sorpresa positiva è significativamente più alta rispetto a quella dei giudizi ottenuti dalla frase che comunica sorpresa negativa. I due enunciati sembrerebbero essere stati, quindi, percepiti dagli ascoltatori con livelli di sorpresa diversi, una tendenza probabilmente connessa alla differenza di valenza dell'emozione convogliata.

Tabella 2 - *Confronto statistico fra i punteggi ottenuti da ciascuno stimolo manipolato nei due enunciati CL_1T e CL_3T*

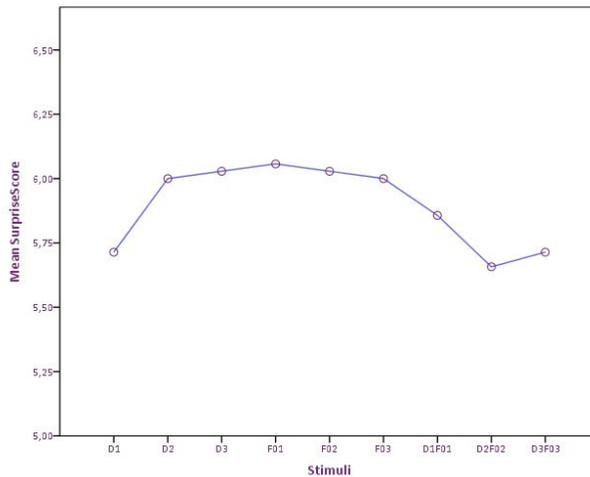
Confronto fra stimoli	Diff. media	t	df	Sig.
D1_1T - D1_3T	1,17	4,797	34	<0,001
D2_1T - D2_3T	1,46	5,166	34	<0,001
D3_1T - D3_3T	1,89	7,294	34	<0,001
F01_1T - F01_3T	1,89	6,393	34	<0,001
F02_1T - F02_3T	1,57	6,039	34	<0,001
F03_1T - F03_3T	1,29	4,410	34	<0,001
D1F01_1T - D1F01_3T	1,37	4,928	34	<0,001
D2F02_1T - D2F02_3T	1,11	3,853	34	<0,001
D3F03_1T - D3F03_3T	1,74	5,259	34	<0,001

Sulla base dei risultati ottenuti da questa analisi preliminare, al fine di comprendere se i giudizi dei parlanti siano influenzati dalla manipolazione dei parametri, i punteggi sono stati sottoposti all'analisi statistica tramite *Friedman test*⁸ separatamente per enunciato. L'analisi ha rivelato che gli stimoli manipolati non differiscono in

⁸ Il *Friedman test* è la versione non parametrica del *one-way ANOVA with repeated measures*. Si tratta di un modello impiegato per testare la differenza fra gruppi nei casi in cui la variabile dipendente sia ordinale. Solitamente, il Friedman test è impiegato anche quando non siano rispettati i presupposti necessari allo svolgimento del modello parametrico corrispondente. Nel nostro caso, la scelta è ricaduta su questo test poiché la variabile in esame è tendenzialmente riconosciuta come ordinale (si tratta dei punteggi ricavati dalla scala Likert utilizzata dai partecipanti al test percettivo).

modo statisticamente significativo in termini di livello di sorpresa percepito⁹, né in riferimento alla frase 1T ($X^2(8) = 9,202, p > 0,05$), né a quella 3T ($X^2(8) = 11,692, p > 0,05$); in altre parole, l'abbassamento dei valori di Onset e l'accorciamento della DVT non hanno comportato una relativa diminuzione del livello di sorpresa percepito. I *line plot* riportati nelle Figg. 9 e 10 aiutano a comprendere le tendenze. Nel caso di *Luca è arrivato!* (Fig. 9) il punteggio più basso di sorpresa percepita è stato attribuito al primo step di manipolazione della durata (D1) e agli stimoli di manipolazione congiunta (D2F02-D3F03). Tale dato non era disatteso: ci si aspettava, infatti, che l'abbassamento congiunto dei valori frequenziali e temporali avrebbe comportato una diminuzione del grado di sorpresa trasmesso. Tuttavia, alcune tendenze sono poco chiare, il primo livello di manipolazione della durata, ad esempio, ha ottenuto un *SurpriseScore* più basso rispetto agli altri stimoli. In particolare, risulta inatteso il fatto che alle variazioni frequenziali non siano corrisposti abbassamenti nei giudizi.

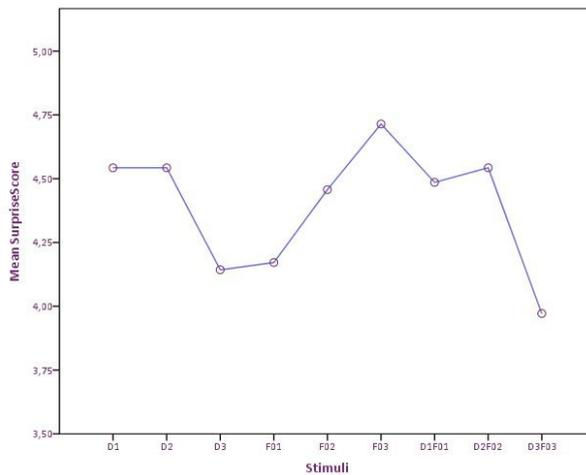
Figura 9 - *Line plot del SurpriseScore attribuito alle sette categorie di stimoli per la frase "Luca è arrivato" (CL_1T)*



Passiamo ad analizzare il *lineplot* dei punteggi raccolti per la frase *Mara è partita!* (Fig. 10).

⁹ Da qui in poi il punteggio attribuito dagli uditori al livello di sorpresa percepito sarà definito *SurpriseScore*.

Figura 10 - Line plot del SurpriseScore attribuito alle sette categorie di stimoli per la frase "Mara è partita!" (CL_3T)



Come è stato anticipato, il confronto non è supportato da significatività statistica, alcune tendenze, però, sono interessanti: a una maggiore riduzione della durata sembra corrispondere un indebolimento del senso di sorpresa trasmesso. Allo stesso modo, un abbassamento interessante si rileva in corrispondenza della manipolazione congiunta D3F03. Il quadro più particolare è quello relativo alla manipolazione dell'Onset di F0: seppure in modo non significativo abbiamo un lieve innalzamento del senso di sorpresa percepito nel passaggio dal primo al terzo test di manipolazione.

8. *Discussione e conclusioni*

L'attenzione finora rivolta alla prosodia delle espressioni di sorpresa si è concentrata in pochi studi dedicati alla codifica e alla decodifica delle emozioni base. Ciò può essere, presumibilmente, ricondotto a due principali motivazioni. In primis, secondo alcuni studiosi la sorpresa tende a essere comunicata principalmente mediante espressioni facciali; è probabile che tale aspetto renda più ostica un'analisi fonetica dei parametri prosodici che caratterizzano questa emozione sia in produzione che in percezione. Inoltre, i confini tra la sorpresa e le altre emozioni sono spesso sfumati: la sorpresa in alcuni contesti può, infatti, sfociare nella gioia o, diversamente, nella rabbia. Si tratta quindi di un'emozione base poco indagata a livello acustico e ancor meno osservata separatamente in relazione alle diverse valenze che la caratterizzano e che in questa sede abbiamo voluto mettere sotto esame, analizzando la modalità frasale che per definizione dovrebbe veicolarla.

Dagli studi condotti fino ad oggi, è emerso che questa emozione sarebbe contraddistinta da alti valori di F0, un'estensione tonale ampia, un'intensità alta, un'eloquio veloce e quindi una riduzione del numero di pause. Si tratta di un quadro prosodico che secondo l'approccio dimensionale farebbe rientrare la sorpresa fra le emozioni

ad alta attivazione. Tuttavia, come sappiamo, con riferimento alla dimensione della valutazione, le emozioni con connotazione positiva si oppongono a emozioni con connotazione negativa. Ci si è chiesti, dunque, se esistano delle differenze fra sorpresa positiva e sorpresa negativa a livello prosodico. L'intento di verificare le caratteristiche intonative delle espressioni di sorpresa si scontra, quindi, con queste premesse e con le numerose variabili che possono presentarsi in un contesto che, come è ovvio, può deviare da una situazione standard. I risultati del presente studio vanno quindi valutati con cautela. Secondo quanto emerge dalla verifica dei dati, la nostra ricerca ha messo in luce che la sorpresa è caratterizzata dall'attivazione dei parametri di frequenza e durata: questa emozione è infatti contraddistinta da valori di F0 più alti e da un'estensione tonale più ampia. In questo caso, il nostro studio si allinea a quanto riscontrato nei contributi che finora hanno dedicato la loro attenzione alla comunicazione vocale della sorpresa (tra gli altri, Cahn, 1990; Abelin, Allwood, 2000; Iriondo et al., 2000; De Marco, Paone, 2014; Maffia et al., 2014). Per quanto concerne il parametro della durata, i nostri dati si discostano dalla tendenza riscontrata in questi contributi, dai quali emerge una sorpresa caratterizzata da un eloquio veloce, per allinearsi con i risultati di Abelin, Allwood (2000) riferiti al tedesco, in cui nella sorpresa si riscontra una durata maggiore rispetto al parlato neutrale. Nella nostra indagine, si è, inoltre, rivolta l'attenzione al confronto fra sorpresa positiva e negativa. In questo caso è emerso che la sorpresa positiva è più intensa e presenta un'estensione tonale più ampia rispetto alla sorpresa negativa. Gli enunciati che comunicano sorpresa positiva sono, quindi, caratterizzati sul piano prosodico da quegli indici fonetici che nelle ricerche finora eseguite sono stati associati all'espressione della sorpresa, tali aspetti sono, però, meno marcati negli enunciati che comunicano sorpresa negativa. Questo quadro ci suggerisce quanto ancora poco si conosca sul livello soprasegmentale di alcune espressioni connesse alle emozioni primarie e quanto la variabilità derivante dal contesto possa influire sulle sfaccettature delle stesse.

Tale dato ha quindi posto l'accento sul fatto che anche per la sorpresa è necessario considerare le diverse sfumature emotive. Scherer (2003) parla di famiglie di emozioni, si distingue ad esempio fra *exuberant joy* e *quite happiness*, oppure fra *hot* e *cold anger*, sostenendo che, in alcuni contesti, i membri di una stessa famiglia emotiva si differenzino significativamente a livello vocale. Anche la sorpresa può, probabilmente, avere varie anime, a seconda della forza con cui l'evento inaspettato colpisce emotivamente il parlante. Oltre a ciò si ha un'ulteriore dicotomia tra sorpresa positiva e sorpresa negativa, declinazioni legate alla portata piacevole o spiacevole dell'evento e della notizia inaspettata che stimola la reazione emotiva. La nostra riflessione, però, prosegue traendo spunto anche da alcune criticità rilevate in altri test percettivi volti all'identificazione e al corretto riconoscimento delle emozioni. Un evento inaspettato suscita sorpresa nel parlante, tuttavia nell'espressione emotiva i confini fra le varie emozioni possono essere sfumati: il soggetto può essere sorpreso ma se l'evento che ha scatenato tale emozione è particolarmente gradito si può sfociare nella gioia o, in caso diverso, nella rabbia. Da questo sfrangiamento dei contorni deriva anche la difficoltà negli studi sperimentali basati sul metodo della *simulated emotional expression*

nel calibrare perfettamente la sceneggiatura che possa stimolare una reazione di pura sorpresa nel parlante. Quindi ci si chiede cosa succede a livello prosodico quando un'emozione è ibrida: poiché è vero che le emozioni ad alta attivazione condividono una prosodia contraddistinta dall'innalzamento di una serie di parametri, ma diversi studi ci suggeriscono che esse si distinguono per alcuni aspetti. Partendo da questo presupposto sarebbe interessante capire cosa accade a livello prosodico quando si innesca uno sconfinamento del perimetro che contorna ciascuna emozione.

Per quanto concerne la percezione, nella nostra ricerca non si è voluto indagare le modalità di riconoscimento o di identificazione delle emozioni. Si è inteso, infatti, osservare la misura in cui alcuni parametri prosodici influenzano il livello di sorpresa percepito dagli uditori. Si è ipotizzato che esistano diversi gradi di sorpresa che un soggetto può provare e si è voluto, quindi, verificare se la manipolazione di alcuni parametri possa influenzare tale percezione determinando una variazione nel livello dell'emozione decodificata.

Il nostro studio ci suggerisce da un lato che l'abbassamento dei valori dell'Onset di F0 e della durata dell'ultima vocale tonica non comportano un indebolimento del grado di sorpresa trasmesso, dall'altro ci mostrano delle tendenze di difficile lettura e che meritano di essere chiarite, probabilmente mediante la ripetizione del test sottoponendolo a un campione più numeroso, valutando, inoltre, altri parametri probabilmente coinvolti in misura maggiore in questo processo.

Un altro aspetto su cui occorrerebbe concentrarsi in futuro sono le modalità di percezione della sorpresa in base alla connotazione affettiva della stessa, quindi sorpresa positiva e negativa. Dai nostri risultati, infatti, i giudizi espressi per i due enunciati selezionati differiscono significativamente: la frase che esprime sorpresa positiva, indifferentemente dal tipo di manipolazione, ha ottenuto punteggi superiori rispetto alla frase che comunica sorpresa negativa; in altri termini, i parlanti hanno percepito una differenza nel livello di sorpresa provato dal parlante. Tale tendenza merita di essere ulteriormente approfondita.

È importante sottolineare, infine, che il disegno sperimentale ha previsto l'analisi di parlato semirecitato. La questione relativa alla modalità di elicitazione degli stimoli più adatta all'analisi della comunicazione delle emozioni è stata a lungo dibattuta (es. Russell, 1994), in particolare, ci si è chiesti se il parlato letto potesse essere rappresentativo dell'espressione spontanea delle emozioni. Secondo Scherer (2003) ogni modalità di elicitazione dello stato emotivo dello speaker può comportare delle problematiche. L'autore suggerisce che la *simulated vocal expression* (che dovrebbe coinvolgere dei parlanti che abbiano delle esperienze attoriali anche minime) sarebbe il metodo di elicitazione delle emozioni più utilizzato in questo filone di studi. Tuttavia, secondo la prospettiva di Scherer non si può escludere che gli attori esagerino volutamente alcuni aspetti, minimizzandone degli altri: in altre parole potrebbero essere influenzati dagli stereotipi legati al parlato emotivo. Ciononostante, il fatto che i *vocal portrayals* siano decodificati correttamente dagli ascoltatori suggerirebbe che il parlato controllato probabilmente rispecchia le caratteristiche acustiche del parlato emotivo naturale. Secondo alcuni autori, il parlato emotivo simulato (*emo-*

tional portayals), basandosi su espressioni naturali, dovrebbe dividerne le caratteristiche acustiche (Banse, Scherer, 1996; Juslin, Laukka, 2003), pensiero sostenuto anche da Pell et al. (2009). Tuttavia, studi più recenti, come quello di Dharmyal, Memon, Raj & Singh (2019), rivelano una differenza significativa tra parlato recitato e parlato spontaneo a livello fonetico, suggerendo di optare per il secondo, in quanto ritenuto più valido e affidabile nello studio della comunicazione emotiva.

Nel caso del presente lavoro si è preferito un parlato letto, basato su delle sceneggiature preconfezionate contenenti le frasi target che, in riferimento al contesto, dovrebbero veicolare l'emozione sorpresa del parlante. Questa scelta è stata effettuata con riferimento alla prospettiva degli studi degli autori citati nonché sulla base dell'intenzione di attuare una comparazione diretta e sistematica tra enunciati identici per struttura sintattica e morfologica e ottenere, dunque, dei dati fonetici e prosodici confrontabili. Nel nostro caso, invitando a partecipare alla sessione di registrazione dei parlanti che non avessero esperienze attoriali, si è inoltre cercato di evitare la possibilità di incorrere nell'eccessiva enfattizzazione della produzione del parlato emotivo.

Tuttavia, consapevoli che l'utilizzo di un parlato ecologicamente valido sia utile e necessario all'indagine delle caratteristiche del parlato emotivo, ci si riserva di spaziare nello studio della sorpresa in questo senso in uno studio futuro.

Riferimenti bibliografici

- ABELIN, Å., ALLWOOD, J. (2000). Cross linguistic interpretation of emotional prosody. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, Newcastle, UK, 5-7 September 2000, 110-113.
- BANSE, R., SCHERER, K.R. (1996). Acoustic profiles in vocal emotion expression. In *Journal of Personality and Social Psychology*, 70(3), 614-636. <https://doi.org/10.1037//0022-3514.70.3.614>
- BELIN, P., FILLION-BILODEAU, S. & GOSSELIN, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for re Behavior Research search on auditory affective processing. In *Behavior research methods*, 40(2), 531-539. <https://doi.org/10.3758/brm.40.2.531>
- BOERSMA, P., WEENINK, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.43, retrieved September, 08 2018 from <http://www.praat.org/>.
- CAHN, J.E. (1990). The generation of affect in synthesized speech. In *Journal of the American Voice I/O Society*, 8, 1-19.
- DARWIN, C. (1872). *The expression of emotions in man and animals*. Londra: John Murray. [Trad. it.: FERRARI G.A. (a cura di) (1982). *L'espressione delle emozioni negli uomini e negli animali*. Torino: Boringhieri].
- De MARCO, A., PAONE, E. (2014). L'espressione e la percezione delle emozioni vocali in apprendenti di Italiano L2. Uno studio cross-linguistico. In *Educazione linguistica. Language education*, 3, 483-500. <http://doi.org/10.14277/2280-6792/339>
- DHAMYAL, H., MEMON, S.A., RAJ, B. & SINGH, R. (2020). The phonetic bases of vocal expressed emotion: natural versus acted. In *ArXiv*. <https://arxiv.org/abs/1911.05733v3>

EKMAN, P. (2003). *Emotions Revealed. Understanding faces and feelings*. London: Weidenfeld and Nicolson.

IRIONDO, I., GUAUS, R., RODRÍGUEZ, A., LÁZARO, P., MONTOYA, N., BLANCO, J.M., BERNADAS, D., OLIVER, J.M., TENA, D. & LONGHI, L. (2000). Validation of an Acoustical Modelling of Emotional Expression in Spanish using Speech Synthesis Techniques. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, Newcastle, UK, 5-7 September 2000, 161-166.

IZARD, C.E. (1994). Innate and Universal Facial Expression: Evidence From Developmental and Cross-Cultural Research. In *Psychological Bulletin*, 115(2), 288-299. <https://doi.org/10.1037/0033-2909.115.2.288>

GILI FIVELA B. (2015). L'integrazione di informazioni multimodali: prosodia ed espressioni del volto nella percezione del parlato. In PISTOLESI E., PUGLIESE R. & GILI FIVELA B. (a cura di), *Parole, gesti, interpretazioni. Studi linguistici per Carla Bazzanella*. Roma: Aracne, 107-127.

JOHNSTONE, T., SCHERER, K.R. (2000). Vocal communication of emotions. In LEWIS, M., HAVILAND, J. (Eds.), *The handbook of emotions*. New York: Guilford, 226-235.

JUSLIN, P.N., LAUKKA, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. In *Psychological Bulletin*, 129(5), 770-814. <https://doi.org/10.1037/0033-2909.129.5.770>

MAFFIA, M., PELLEGRINO, E. & PETTORINO, M. (2014). Labelling expressive speech in L2 Italian: The role of prosody in auto-and external annotation. In CAMPBELL, A.W., GIBBON, D. & HIRST, D. (Eds.), *Proceedings of the 7th International Conference on Speech Prosody*, Dublin, Ireland, 20-23 May 2014, 81-84.

PELL, M.D., PAULMANN, S., DARÁ, C., ALASSERI, A. & KOTZ, S.A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. In *Journal of Phonetics*, 37, 417-435. <https://doi.org/10.1016/j.jwocn.2009.07.005>

RUSSELL, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. In *Psychological Bulletin*, 115(1), 102-41. <https://doi.org/10.1037/0033-2909.115.1.102>

SCHERER, K.R. (1989). Vocal correlates of emotional arousal and affective disturbance. In WAGNER, H., MANSTEAD, A. (Eds.), *Handbook of Psychophysiology: Emotion and social behavior*. London: Wiley, 165-197.

SCHERER, K.R. (2000). Psychological models of emotion. In BOROD, J. (Ed.), *The Neuropsychology of Emotion*. Oxford: Oxford University Press, 137-162.

SCHERER, K.R. (2003). Vocal communication of emotion: a review of research paradigms. In *Speech communication*, 40 (1-2), 227-256. [https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5)

TOMKINS, S. (1962). *Affect Imagery Consciousness: Volume I, The Positive Affects*. London: Tavistock. <https://doi.org/10.1037/14351-000>

TOMKINS, S. (1963). *Affect Imagery Consciousness: Volume II, The Negative Affects*. London: Tavistock.

WUNDT, W. (1874/1905). *Grundzüge der physiologischen Psychologie. Fundamentals of physiological psychology*, orig. pub. 1874 fifth ed. Engelmann, Leipzig.