

# LA COMUNICAZIONE DELLE EMOZIONI ATTRAVERSO IL CANALE VOCALE NON VERBALE

## 1. Introduzione

Il capitolo precedente illustra il ruolo fondamentale dell'espressione del viso nel permetterci di comunicare, recepire e condividere le emozioni, traducendo una sensazione che altrimenti rimarrebbe un fatto privato in un efficace mezzo di comunicazione di scambio interpersonale. Come anticipato tuttavia, le espressioni facciali non sono l'unico mezzo che abbiamo a disposizione. La voce è infatti uno strumento potente, che permette di esprimere e di cogliere l'infinita gamma delle sfumature emotive.

## 2. Gli indicatori vocali implicati

La voce comunica le emozioni soprattutto attraverso i parametri dell'intonazione, dell'intensità e del ritmo che, nella classificazione delle caratteristiche paralinguistiche, costituiscono il *sistema prosodico o soprasegmentale* (Anolli & Cicceri, 1997). Innanzitutto, ciò che noi percepiamo come *tonalità* è dato dalla frequenza fondamentale ( $F_0$ ), che è determinata dalla tensione e dalla vibrazione delle corde vocali e dell'intera laringe. Più esse sono tese, più la frequenza è alta e più la tonalità è acuta; al contrario, viene percepito un tono grave quando esse sono allentate. La tonalità viene misurata in Hertz. Più nello specifico, le variabili relative al tono sono le seguenti:

- media di F0 (frequenza fondamentale): è la media aritmetica di F0 in un enunciato;
- gamma di F0: è la differenza tra il tono di voce più alto e quello più basso all'interno di un enunciato;
- variazione di F0: si ottiene contando il numero di variazioni (picchi e deflessioni) dell'intonazione all'interno dell'enunciato;
- profilo di intonazione: andamento dell'enunciato, ottenuto congiungendo tutti i punti indicativi della tonalità.

Rispetto ai *parametri temporali*, vengono distinti la durata, la velocità di emissione e il ritmo. La durata consiste nella lunghezza dell'enunciato misurata in secondi. La velocità di emissione è invece data dal numero di sillabe pronunciate in un secondo. Viene operata un'ulteriore distinzione tra la velocità dell'eloquio (velocità di emissione dell'enunciato comprendente suoni e pause) e velocità di articolazione (velocità di emissione dei soli periodi di suono, escluse le pause). Infine, il ritmo consiste nella combinazione tra durata e accento ed è dato dall'alternarsi di sillabe accentate e sillabe non accentate.

Da ultimo, ciò che noi soggettivamente percepiamo come volume è l'*intensità*, dovuta alla pressione della laringe e alla forza fonoespiatoria. L'intensità viene misurata in decibel (db). Perché un suono venga percepito è necessario che la sorgente sonora emetta vibrazioni di frequenza comprese tra i 20 e i 20.000 cicli al secondo. Al di sotto (infrasuoni) e al di sopra (ultrasuoni) di tali soglie il suono non viene percepito dall'orecchio umano. Rispetto all'intensità vengono considerati i seguenti parametri:

- media: è la media aritmetica dell'intensità in un enunciato;
- gamma: è la differenza tra l'intensità maggiore e quella minore all'interno di un enunciato;
- profilo di intensità: andamento dell'enunciato, ottenuto congiungendo tutti i punti indicativi dell'intensità.

### 3. Gli studi empirici sulla comunicazione vocale delle emozioni

La comunicazione vocale delle emozioni è stata esplorata focalizzando l'attenzione su due diversi processi:

- le ricerche sul processo di encoding indagano le modalità di codifica vocale delle espressioni emotive;
- gli studi relativi al processo di decoding cercano di capire come avvenga il riconoscimento delle emozioni a partire dalle caratteristiche vocali non verbali.

#### 3.1. Problemi metodologici

Nel loro insieme, le ricerche si sono dovute confrontare con diversi problemi di ordine metodologico (Davidson, Scherer & Goldsmith, 2003). Una prima difficoltà riguarda la *scelta dei parametri* da utilizzare tra i diversi possibili (vedi par. 3.2.). In secondo luogo, i ricercatori devono stabilire se utilizzare *enunciati spontanei*, o *indotti sperimentale* oppure *simulati*. A tale proposito, da tempo sono stati messi in evidenza i vantaggi della simulazione, che permette una maggiore possibilità di controllo e consente di attribuire un carattere di enfasi e ridondanza dell'encoding, riducendo così l'ambiguità. Infatti, l'utilizzo di stimoli raccolti in una situazione in cui realmente le persone esprimono le emozioni attraverso la voce costituirebbe una condizione ideale. La spontaneità ha però come svantaggio la mancanza di controllo. Per questa ragione, anche se ci sono diverse ragioni per sospettare l'esistenza di differenze tra l'*emotional speech* spontaneo e quello recitato, solitamente viene chiesto a dei parlanti - meglio se si tratta di attori professionisti - di simulare le diverse emozioni. Maggiormente naturale, anche se di più difficile realizzazione, è la registrazione di materiale che deriva dall'induzione in laboratorio di una reale emozione nel parlante.

Un ulteriore problema metodologico riguarda la scelta delle *tecniche di mascheramento*, le quali permettono di mantenere inalterate le caratteristiche paralinguistiche eliminando però il contenuto semantico verbale. Esistono diversi tipi di mascheramento:

- utilizzo di suoni, sillabe senza senso o liste di lettere o numeri;
- filtraggio elettronico, che consente di alterare l'informazione verbale, rendendola incomprensibile. Ad esempio, il filtraggio della frequenza fondamentale consente di limitarne la gamma, attraverso filtri passa-basso, che permettono il passaggio delle sole frequenze basse, o attraverso filtri passa-alto, che permettono al contrario il passaggio delle sole frequenze alte. In particolare i filtri passa-basso, eliminando le frequenze alte, che sono indispensabili per il riconoscimento del parlato, rendono inintelligibile il contenuto verbale. Tuttavia, bisogna osservare se l'ascolto viene ripetutamente reiterato, l'ascoltatore può giungere alla comprensione del contenuto verbale;
- capovolgimento o taglio e rimontaggio del nastro;
- utilizzo di una lingua straniera;
- utilizzo di frasi neutre, cioè prive di un contenuto verbale emotivamente connotato. Quest'ultima tecnica di mascheramento presenta evidenti vantaggi, in quanto preserva un profilo paralinguistico che va oltre la singola sillaba o parola, mantenendo inoltre invariati i parametri soprasegmentali.

Recentemente, Douglas-Cowie e colleghi (Douglas-Cowie *et al.*, 2003) hanno passato in rassegna i database esistenti relativi all'*emotional speech* (emozioni espresse attraverso il canale vocale) delineando alcune caratteristiche che dovrebbero essere tenute presenti nella costruzione di un nuovo database. Come essi sottolineano, bisogna innanzitutto tenere in considerazione il fatto che il modo di esprimere vocalmente un'emozione varia considerevolmente da parlante a parlante, da cultura a cultura, tra i generi e tra le condizioni socio-ambientali. Di conseguenza, per garantire che i risultati di una ricerca volta ad indagare l'*emotional speech* siano generalizzabili è necessario, nella costruzione della batteria degli stimoli uditivi, considerare una serie di fattori. In primo luogo, bisogna valutare attentamente il

numero e il tipo di parlanti coinvolto nella costruzione del campione. Gli autori notano che la maggior parte degli studi ha utilizzato un unico soggetto. Per ottenere risultati altamente generalizzabili bisognerebbe coinvolgere diversi parlanti, diversificati sulla base della lingua e del genere. E' auspicabile che i parlanti siano attori professionisti, poiché essi presentano una maggiore competenza nell'utilizzare i parametri vocali in modo intenzionale. Sempre allo scopo di permettere la generalizzazione dei risultati, è bene utilizzare un'ampia gamma di emozioni. La rassegna effettuata indica tuttavia che di norma viene impiegato un numero ridotto di emozioni, per lo più rabbia, tristezza, paura, stress e gioia. Un altro dato che emerge dalla rassegna è che solitamente il materiale è avulso da qualsiasi contesto: si tratta infatti per lo più di registrazioni audio di brevi enunciazioni con contenuto semantico neutro e struttura sintattica prestabilita. A tale proposito bisogna fare alcune considerazioni: nella situazione naturale le persone solitamente utilizzano diversi *indizi contestuali* per integrare le informazioni convogliate dal vocale al fine dell'identificazione dell'emozione. Possiamo indicare almeno quattro diversi tipi di contesto, che dovrebbero essere tenuti in considerazione: un contesto semantico, relativo al contenuto verbale; un contesto strutturale, costituito dalla struttura sintattica della frase; un contesto intermodale, poiché le informazioni acustiche sono integrate da quelle offerte dalle altre modalità sensoriali, principalmente quella visiva; un contesto temporale, dal momento che l'emozione non è espressa in modo puntuale e può essere definita come un processo. Infine, è utile osservare che i parametri più frequentemente utilizzati per analizzare il materiale sono l'intensità, l'intonazione, la struttura formale dello spettro e i parametri temporali (Banse & Scherer, 1996).

### *3.2. Le ricerche sull'encoding*

Le ricerche sulla codifica vocale delle emozioni hanno l'obiettivo di stabilire quali cambiamenti nella voce e nella produzione del parlato - e quali corrispondenti patterns di parametri acustici - si manifestano come conseguenza dell'insorgere di

un'emozione nel parlante. Più nello specifico, tali studi intendono far luce sull'esistenza di differenze nell'espressione vocale delle diverse emozioni. E' possibile operare una rudimentale distinzione tra le differenti emozioni sulla base dei parametri acustici. Tuttavia, il legame tra le caratteristiche prosodiche e la specifica emozione è abbastanza approssimativo, dal momento che esiste un largo margine di variabilità, sia nell'espressione e nel riconoscimento delle emozioni, sia a livello di variabilità intra-individuale e inter-individuale. Aubergé e Cathiard (2003) hanno ad esempio rilevato che le differenze individuali giocano un ruolo non secondario nell'espressione e nel riconoscimento del divertimento, confermando l'intuizione dell'esistenza di diverse strategie per esprimere vocalmente la stessa emozione.

Bisogna inoltre tenere in massima considerazione il fatto che, come precedentemente accennato (vedi par. 3.1), il significato delle caratteristiche acustiche è contesto-dipendente (Cook, 2002).

Come tendenza generale, gli studi che hanno indagato l'encoding delle emozioni a livello vocale non verbale hanno evidenziato l'esistenza di due pattern: l'uno, caratterizzato da alta frequenza fondamentale, ampia gamma e variabilità della frequenza fondamentale, elevata intensità e alta velocità, è tipico delle emozioni ad alto livello di arousal (livello di attivazione fisiologica); l'altro, identificato da bassa frequenza fondamentale, gamma e variabilità di intonazione ristrette, debole intensità e velocità ridotta, è invece tipico delle emozioni a basso arousal (Banse & Scherer, 1996). In effetti, i parametri vocali sono un potente indicatore del livello di attivazione fisiologica, come suggeriscono le cosiddette "teorie dell'arousal" (Pittam & Scherer, 1993). Tuttavia, esistono diverse evidenze, messe in evidenza dalle "teorie dell'appraisal", del fatto che in realtà l'espressione vocale delle emozioni è determinata dall'azione congiunta dell'arousal e della *valenza edonica* : in effetti, le modalità specifiche di encoding di alcune emozioni non possono essere spiegate facendo riferimento esclusivamente al livello di attivazione (Klasmeyer & Sendlmeier, 1997). Di conseguenza, un modello bidimensionale, se non *multidimensionale* dell'encoding emotivo appare oggi più plausibile di un modello unidimensionale. Un simile modello è inoltre supportato da studi recenti che hanno

nesso in evidenza l'esistenza di differenze qualitative nell'espressione delle diverse emozioni (Juslin & Laukka, 2001; Scherer, Johnstone & Klasmeyer, 2003). Nello specifico, Anolli e Ciceri (1997) hanno delineato la "voce" di alcune emozioni. Ad esempio, la gioia è caratterizzata da una media della frequenza fondamentale molto elevata (media di  $F_0$  pari a 228 Hz) e quindi da una voce acuta; elevate sono anche la gamma e la variazione dell'intonazione, con un profilo di intonazione variato e mosso, ma con variazioni sempre graduali e lente, e quindi arrotondate. Il profilo discende gradualmente nella parte finale dell'enunciato per poi assestarsi su una tonalità intermedia o per risalire lievemente. Anche l'intensità è elevata e ha un profilo ascendente nella parte conclusiva dell'enunciato. La velocità di eloquio e di articolazione è media, la durata è standard e le pause presentano valori medi. La *tristezza* ha un profilo di intonazione caratterizzato da una frequenza fondamentale bassa (media di  $F_0$  pari a 134 Hz), con variazioni quasi inesistenti e arrotondate e gamma ristretta. Il finale è discendente. Anche il volume e la velocità sono bassi. La tristezza si connota soprattutto per le sue lunghe pause. Nel caso della *paura*, il tono è molto acuto: l'energia si concentra nelle zone di alta frequenza dello spettro ( $F_0$  media pari a 278 Hz), con ampia gamma e numerose variazioni piuttosto brusche, che conferiscono al profilo un andamento spigoloso. L'intensità è elevata e progressivamente crescente: è possibile che l'eloquio giunga addirittura ad essere gridato. Pur avendo una durata della frase standard e una velocità dell'eloquio media, la paura presenta una velocità di articolazione alta. Le pause costituiscono frequenti e rapide "rotture" del parlato, dovute alla respirazione superficiale. La *rabbia* è caratterizzata da una frequenza fondamentale media ( $F_0$  media pari a 178 Hz, inferiore a quella di gioia e paura), con numerose e a volte repentine, spigolose e ampie variazioni del tono. Il profilo ha carattere ascendente. I livelli di intensità sono costantemente alti. Il ritmo è serrato, con una durata della frase contenuta. Si manifesta la tendenza ad "espellere" la frase, in un'unica emissione di respiro e senza pause. Il *disprezzo* si caratterizza per l'articolazione molto lenta delle sillabe e per la durata molto lunga dell'enunciato. Il tono è grave, con una gamma ristretta e variazioni minime e di forma arrotondata. Il profilo presenta un finale discendente.

L'intensità è bassa e costante. Infine, la *tenerezza* presenta un ritmo regolare e costante. Tratto distintivo di questa emozione è la tonalità molto grave, con scarse variazioni di frequenza fondamentale e gamma ristretta. Il volume è basso e decrescente.

Un interessante contributo giunge da uno studio (Gobl & Chasaide, 2003) che ha indagato il ruolo delle *qualità vocali* - ad esempio la voce piena, la voce sottile, la voce ruvida - nel veicolare informazioni circa emozioni, umori e attitudini. A differenza dei parametri di picco, di volume e temporali, che sono relativamente più facili da misurare e da controllare, le qualità vocali sono state finora poco esplorate in relazione alla comunicazione dei correlati emotivi, proprio a causa delle difficoltà metodologiche e concettuali che esse implicano (Scherer, 1986). I risultati indicano che effettivamente le qualità vocali contribuiscono alla comunicazione vocale delle emozioni, non tanto in forza di una corrispondenza biunivoca tra qualità della voce e tipo di emozione espressa ma piuttosto in virtù di un'associazione tra una certa qualità vocale e una costellazione di stati affettivi. Inoltre, la qualità vocale risulta più efficace nel comunicare la valenza edonica rispetto che l'arousal, che viene invece meglio veicolato da intensità, tonalità e durata. Gli autori ipotizzano che la frequenza e la qualità vocale contribuiscano in modo indipendente all'espressione degli stati affettivi: le qualità vocali hanno un ruolo di primo piano nel comunicare gli stati affettivi meno intensi come stati d'animo, umori e attitudini, soprattutto se privi di un correlato fisiologico, mentre la frequenza è fondamentale per l'espressione delle emozioni più intense, accompagnate da una componente a livello fisiologico. La possibilità che la frequenza e le qualità vocali abbiano funzioni differenti e potenzialmente indipendenti nell'espressione degli affetti è stata precedentemente suggerita anche da Scherer (1986), il quale ha ipotizzato che le ampie variazioni dell'intonazione segnalino marcate differenze nei livelli di attivazione, e che invece le qualità vocali comunichino sottili sfumature affettive.

### 3.2.1. Il modello di Scherer

Scherer (1986) individua una relazione tra i *parametri vocali* delle singole emozioni e le componenti di *valutazione cognitiva* (*appraisal*) delle circostanze attivanti contemplate dal *modello componenziale* da lui stesso proposto. Tale modello individua cinque livelli di valutazione dello stimolo:

- *novità*: il carattere di novità di uno stimolo inatteso provoca una risposta di orientamento, dovuta alla rilevazione di una discrepanza tra le aspettative e lo stimolo. Stimoli poco familiari suscitano reazioni di sorpresa, di curiosità e di avvicinamento, ma anche di paura e diffidenza. Mentre le emozioni positive sono suscitate dalla coincidenza tra le aspettative e i desideri dell'individuo e lo stimolo emotigeno, mentre quelle negative sono il risultato di una discrepanza tra di essi;
- *valenza edonica*: consiste nel carattere di piacevolezza/spiacevolezza dello stimolo. Nel caso in cui esso sia valutato come piacevole, vengono attuate delle risposte di avvicinamento, inclusione e gioia, in caso contrario invece si riscontrano condotte di esitamento, fuga e attacco;
- *rilevanza rispetto ai bisogni e agli scopi del soggetto*: il fatto che uno stimolo costituisca la risposta ad uno scopo, desiderio o bisogno del soggetto, contribuisce ad attribuirgli rilevanza e a suscitare una reazione di avvicinamento, ed in particolar modo di gioia. In caso contrario, la reazione abituale è la collera;
- *capacità di far fronte (coping) allo stimolo*: l'individuo attua due test: tramite il test di controllo, che ha come oggetto le cause elicитanti, e tramite il test di potenza, che ha come oggetto le risorse personali, egli verifica se è possibile controllare o annullare l'evento che suscita la risposta emozionale. La consapevolezza dell'efficacia delle proprie risorse suscita un senso di padronanza, mentre la consapevolezza della propria impotenza genera depressione e ansia;
- *compatibilità con l'immagine di sé e le norme sociali*: l'adesione o la trasgressione delle regole sociali e dei valori personali suscitano differenti reazioni emozionali. A tale proposito, particolare rilievo va attribuito alle cosiddette emozioni sociali

o secondarie, come la vergogna, l'imbarazzo, la colpa e l'orgoglio, che emergono nel processo di socializzazione. Tali emozioni costituiscono un feed-back della propria condotta, ricoprendo così un ruolo di monitoraggio.

Scherer suggerisce che le componenti di valutazione citate producano specifici effetti a livello del sistema nervoso centrale, i quali si traducono a loro volta in modificazioni della tensione della muscolatura dell'apparato fonatorio. In particolare, secondo Scherer, gli *effetti tonici* -più durevoli- non sarebbero controllati volontariamente, a differenza degli *effetti fasici* di rilassamento - a carattere transitorio -, che sarebbero invece il prodotto dei tentativi dell'organismo di controllare l'espressione vocale emotiva. Anche i cambiamenti che avvengono nel sistema nervoso autonomo, quali la modificazione del ritmo respiratorio e della secrezione di saliva, hanno degli effetti sull'emissione vocale.

Scherer infine distingue tra voce piena e voce sottile. La *voce piena*, caratterizzata da una modesta tensione e una respirazione piena, si manifesta quando l'individuo ha fiducia nelle proprie risorse e si prepara all'attacco. La *voce sottile* invece, che risulta da una respirazione poco profonda e una aumento della frequenza fondamentale, contraddistingue quelle situazioni in cui l'individuo, valutando la propria impotenza a contrattaccare, si prepara alla fuga.

### 3.3. Le ricerche sul decoding

Le ricerche sulla decodifica vocale delle emozioni, che hanno avuto inizio tra gli anni '50 e gli anni '80, sono volte a comprendere quali variazioni dei parametri acustici ci permettano di cogliere la presenza delle emozioni nella voce dell'altro e di discriminarle. In genere, a questo scopo, una serie di stimoli vocali emotivi prodotta da attori professionisti viene sottoposta ad un campione di giudici, cui viene proposto un compito di discriminazione (scelta di una tra le diverse alternative date) o di riconoscimento (*free labelling*, cioè libero attribuzione di un'etichetta emotiva) (Scherer, Johnstone & Klasmeyer, 2003). Alcune ricerche hanno inoltre manipolato i parametri paralinguistici con lo scopo di evidenziare quali caratteristiche abbiano un

maggior peso ai fini del riconoscimento. I principali risultati possono essere così riassunti: le diverse caratteristiche acustiche hanno una diversa incidenza rispetto alle diverse emozioni; in generale, tuttavia, si può affermare che la *frequenza fondamentale* e l'*intensità*, nonché le loro variazioni (principalmente media e gamma), così come anche il *ritmo di articolazione*, comunicano le informazioni essenziali per il riconoscimento delle emozioni.

I dati sperimentali indicano che il riconoscimento delle emozioni sulla base dei segnali vocali non verbali avviene in percentuali cinque o sei volte superiori al caso: è stata infatti evidenziata un'accuratezza media del 60%. Confrontando questo dato con quello relativo all'accuratezza riconoscimento delle emozioni basato sull'espressione facciale, emerge una maggiore difficoltà nel decoding vocale. La differenza è dovuta in particolar modo al riconoscimento della gioia e del disgusto: queste emozioni sono scarsamente discriminate in base ai parametri vocali, mentre invece sulla base della mimica ottengono valori vicini al 100% (Ekman, 1994). La tristezza e la rabbia - ma anche la paura - sono invece meglio riconosciute a livello vocale. Johnstone e Scherer (2000) hanno attribuito a queste differenze un'origine di tipo evolutivo: è utile che un segnale di rabbia o di paura possa essere comunicato anche a grande distanza. Al contrario, un'impressione di disgusto per aver assaggiato del cibo che non è stato apprezzato viene più facilmente espressa ad un cospecifico che si trova nelle vicinanze.

In generale, dalle ricerche emerge che a livello vocale le emozioni negative sono riconosciute più facilmente e con un'accuratezza maggiore rispetto alle emozioni positive. Questo deriva probabilmente dal valore adattivo che il loro riconoscimento ricopre ai fini della salvaguardia dell'individuo (Balconi & Lucchiari, 2005; Balconi & Pozzoli, 2003). Le emozioni più riconosciute sono la rabbia e il dolore, mentre quelle identificate con maggiore difficoltà sono la gioia, il disgusto e il disprezzo. Le emozioni non vengono confuse casualmente, ma in modo sistematico. Ad esempio, tendenzialmente la tristezza viene confusa con l'amore, la vergogna con la tristezza, la paura con la sorpresa. Inoltre, il disprezzo da un lato e il disgusto e la collera dall'altro vengono frequentemente scambiate; lo stesso accade nel caso di gioia e

sorpresa. Il fatto che la sorpresa venga spesso confusa viene talvolta spiegato facendo riferimento alla presenza di un'iniziale elemento di novità/sorpresa dello stimolo, che in un secondo momento della valutazione cognitiva assume una più chiara connotazione di piacevolezza/piacevolezza. La sorpresa è infatti un'emozione ambivalente, che può assumere una valenza positiva o negativa a seconda della specifica natura dello stimolo elicitante. Sono state avanzate due ipotesi, entrambe plausibili e complementari, per spiegare questi errori sistematici. La prima ipotesi fa riferimento alla condivisione del medesimo livello di arousal, la seconda invece, più competente, alla vicinanza semantica (Russell & Fernández-Dols, 1998) tra le emozioni che vengono confuse. Recentemente, una ricerca condotta da Scherer (1994) ha contribuito a spiegare il motivo per cui alcune le caratteristiche soprasegmentali sono un mezzo efficace per esprimere alcune emozioni ma non altre. Scherer introduce il concetto di *affective bursts*, descrivendoli come espressioni non verbali molto brevi e discrete di un'emozione, che si manifestano a livello del volto e della voce e che sono elicitate da eventi chiaramente identificabili. Se si considera l'aspetto funzionale, il concetto di *affective burst* in parte si sovrappone a quello di "interiezione affettiva". Tuttavia i confini dei due costrutti non coincidono: un'esclamazione verbale come "Cielo!", proprio in virtù della sua natura verbale viene classificato come interiezione, a differenza di una risata, che viene invece identificata come *affect burst*.

Scherer offre una classificazione degli *affective bursts* basata sulla sua distinzione push-pull (Scherer, 1988). Gli *effetti push* sono fattori fisiologici, come il dolore, che si traducono in un'espressione. Gli *effetti pull* invece sono regole e aspettative sociali che prevedono espressioni socialmente appropriate per le diverse situazioni, andando così a modulare le espressioni elicitate dagli effetti push. Entrambi i fattori quasi sempre contribuiscono alla produzione dell'espressione emozionale, ma uno di essi può prevalere sull'altro in date situazioni. Sulla base di tale distinzione, Scherer propone una differenziazione tra i *raw affect bursts* (*affect bursts* grezzi), che si collocano al polo push del continuum e gli *affect emblems* (emblem affettivi), che si collocano invece all'estremità pull del continuum. Di

conseguenza, i raw effect bursts possono essere definiti come vocalizzazioni grezze tendenzialmente universali e minimamente convenzionalizzate, che presentano forti differenze interindividuali. Al contrario gli affect emblems sono simboli convenzionalizzati, strettamente cultura-dipendenti, che presentano scarse differenze individuali. Ovviamente, tra i due estremi del continuum esistono infinite possibilità di bursts misti, che vanno da suoni non articolati (risata) fino a interazioni con una struttura fonemica se pur non verbali (“wow”) (vedi fig. 1).

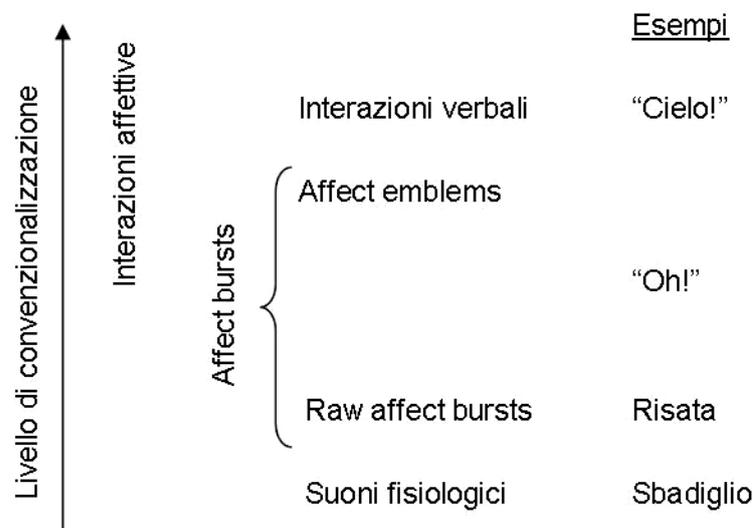


Fig 1. Le interiezioni affettive verbali e gli affective bursts come estremi di un continuum delle espressioni vocali emotivamente connotate (tratto da Schröder, 2003).

In un recente articolo, Schröder (2003) ha dimostrato che gli affective bursts sono in grado, se presentati nella sola modalità uditiva e in assenza di un contesto, di comunicare un significato emotivo chiaramente identificabile. Tutte e 10 le emozioni espresse (ammirazione, paura, disgusto, contentezza, noia, calma, sorpresa, preoccupazione, disprezzo, rabbia) presentano alti indici di riconoscimento (in media 81%). Questo indica che gli affective bursts sono degli efficaci strumenti per l'espressione delle emozioni. Le percentuali di riconoscimento sono notevolmente più alte di quelle rilevate nel caso di decoding basato sulle caratteristiche prosodiche

o sulle qualità vocali del parlato. Ciò è dovuto al fatto che, nel caso degli affective bursts, anche la struttura segmentale stessa, e non solo gli aspetti sovrasegmentali, comunicano un significato. Più nello specifico, si può ipotizzare che alcune emozioni siano più efficacemente espresse dagli affect bursts e altre invece dalla prosodia e dalla qualità della voce. Il disgusto è un chiaro esempio di emozione che viene tipicamente espressa attraverso affect bursts ("bleah"). Banse e Scherer (1996) hanno trovato che gli indici di riconoscimento del disgusto a partire dalla prosodia e dalla qualità vocale del parlato si attestano solo intorno al 15%. Questa emozione infatti viene tipicamente espressa tramite affective bursts anziché attraverso lunghi e articolati enunciati, come supportato dai risultati di questo studio. Al contrario, sembra che altre emozioni, ad esempio la rabbia, che in questo studio ha riportato livelli minori di riconoscimento, siano scarsamente espresse ricorrendo agli affective bursts. In effetti, Johnstone (Johnstone & Scherer, 2000) ha suggerito che alcune emozioni, tra cui la rabbia, vengano espresse tramite una configurazione prototipica di variabili acustiche legate al parlato, e che di conseguenza vengano facilmente prodotte dagli attori e facilmente riconosciute dagli ascoltatori, mentre altre emozioni, come il disgusto, che non presentano questa configurazione tipica, vengano molto più difficilmente prodotte e comprese.

Infine, è emerso che gli affect bursts in generale non sono specifici di una singola emozione ma sono adatti ad esprimere una certa gamma di emozioni, simili se considerate in riferimento alle dimensioni di valenza, arousal e controllabilità.

#### **4. Il contributo della neuropsicologia**

La neuropsicologia, ed in particolar modo gli studi sui deficit comunicativi risultanti da lesioni cerebrali, hanno contribuito a chiarire come avviene il processamento delle informazioni emotive espresse dalla voce.

Rispetto all'encoding, nel complesso, i dati disponibili indicano che alcuni specifici nuclei del tronco encefalico, attivati da proiezioni provenienti dal sistema

limbico, coordinano l'attivazione dei muscoli laringei e respiratori, producendo in questo modo specifici pattern vocali (Scherer et al., 2003).

Maggiore attenzione è stata rivolta ai meccanismi cerebrali sottesi al decoding delle emozioni espresse attraverso i parametri vocali. Se da un lato è universalmente riconosciuto che le componenti linguistiche e grammaticali del parlato vengono elaborate in specifici centri dell'emisfero sinistro, d'altro canto non esiste invece un consenso circa la localizzazione dei centri preposti alla decodifica delle componenti emotive non linguistiche del parlato, sempre ammesso che questi esistano. A partire dagli anni '70, diversi studi hanno portato ad ipotizzare una specializzazione dell'emisfero destro per il processamento delle componenti emotive vocali (Tucker, Watson & Heilman, 1977), confermata dall'osservazione di un deficit nella percezione della prosodia emotiva in pazienti con danno cerebrale unilaterale all'emisfero destro ma non in pazienti con danno unilaterale all'emisfero sinistro (Peper & Irle, 1977). Tuttavia, non tutte le ricerche effettuate hanno riscontrato tale lateralizzazione destra. Van Lancker e Sidtis (1992) hanno suggerito che l'emisfero destro sia deputato non tanto all'elaborazione della prosodia emotiva ma più in generale al processamento di alcuni tipi di informazioni acustiche come il livello e la variabilità della frequenza fondamentale, che sono indici fondamentali per il riconoscimento delle emozioni espresse mediante la voce.

Un importante contributo ai fini della localizzazione di questi processi può giungere dalle tecniche di neuroimaging. Tuttavia, gli studi di questo tipo sono poco numerosi. Un interessante studio condotto con la tomografia ad emissione di positroni (PET) ha riportato una notevole attività prefrontale destra durante il processamento del contenuto proposizionale emotivo di alcune frasi (George et al., 1996). Imaizumi (Imaizumi et al., 1998) ha rilevato, utilizzando la magnetoencefalografia (MEG), sia rispetto al processamento linguistico sia rispetto a quello prosodico, l'esistenza di processi specifici di elaborazione delle differenti componenti prosodiche localizzati nell'emisfero destro ma che si avvalgono anche del contributo di alcuni centri dell'emisfero sinistro.

Un apporto giunge anche dagli studi elettroencefalografici. Tra di essi, una ricerca condotta registrando l'elettroencefalogramma ha riportato una specializzazione della corteccia destra nella percezione dei parametri temporali (durata delle sillabe accentate) e intonazionali (gamma di F0). Un recente studio (Alter *et al.*, 2003) ha inoltre messo in evidenza l'esistenza di differenze legate alle caratteristiche della specifica emozione. Alter ha infatti messo in relazione alcuni parametri legati all'encoding del parlato emotivamente connotato (ad esempio il tipo di respirazione o l'asprezza) con le caratteristiche del tracciato ERP registrato su un soggetto mentre ascoltava. Gli autori hanno individuato, in entrambi i casi, una differenziazione basata sulla valenza edonica, positiva, negativa o neutra.

Infine, bisogna osservare una carenza di studi che abbiano utilizzato la tecnica della registrazione dei potenziali evocati, che è invece stata efficacemente utilizzata nello studio del decoding emotivo del volto.