

CONSERVATORIO DI MUSICA “CESARE POLLINI” DI
PADOVA DIPARTIMENTO DI NUOVE TECNOLOGIE E
LINGUAGGI MUSICALI DIPLOMA ACCADEMICO DI
2° LIVELLO IN MUSICA ELETTRONICA
(TECNICHE INFORMATICHE DI ANALISI E VALORIZZAZIONE
DEI MATERIALI MUSICALI)



ANNO ACCADEMICO 2016-2017

**PERCEZIONE CROSS-MODALE:
DAGLI STIMOLI UDITIVI ALLE
CORRISPONDENZE VISIVE**

studente NICOLA SANGUIN
matricola 12545
relatore prof. NICOLA BERNARDINI
correlatore prof. ALBERTO NOVELLO



Distribution of this text is allowed under
**Creative Commons Attribution-Noncommercial-Share
Alike 3.0 license.**

Please read details on
<https://creativecommons.org/licenses/by-nc-sa/3.0/>

**Percezione Cross-Modale:
dagli Stimoli Uditivi alle Corrispondenze Visive**

Nicola Sanguin

7 Maggio 2018

Indice

| | |
|---|----|
| Ringraziamenti..... | 1 |
| Introduzione..... | 3 |
| Percezione cross-modale: <i>prior art</i> | 7 |
| 1.1 Il mito sinestesico..... | 9 |
| 1.2 Percezione cross-modale..... | 13 |
| 1.3 Metodologie sperimentalistiche per lo studio della percezione cross-modale..... | 19 |
| 1.4 Tassonomia delle corrispondenze crossmodali uditive-visibility..... | 23 |
| 1.4.1 Pitch – Luminosità..... | 26 |
| 1.4.2 Pitch – Dimensione delle forme..... | 28 |
| 1.4.3 Pitch – Elevazione spaziale delle forme..... | 30 |
| 1.4.4 Pitch – Geometria delle forme..... | 33 |
| 1.4.5 Loudness – Luminosità..... | 36 |
| 1.4.6 Loudness – Dimensione delle forme..... | 38 |
| 4.7 Timbro – Geometria delle forme..... | 40 |
| 1.5 Illusioni percettive..... | 43 |
| 1.6 Corrispondenze empiriche suono – colore..... | 46 |
| Percezione cross-modale: bibliografia annotata, 2016-2018..... | 53 |
| 2.1 2016..... | 55 |
| 2.2 2017..... | 57 |
| 2.3 2018..... | 75 |
| Conclusioni..... | 79 |
| Bibliografia..... | 83 |
| Elenco delle figure..... | 99 |

Ringraziamenti

Un affettuoso ringraziamento va alla mia famiglia al completo, per il sostegno che non è mai venuto a mancarmi, soprattutto nei momenti più bui, e per la pazienza che è quotidianamente costretta a portare per sopportare il mio modo di essere non certo facile, a volte rumoroso, spesso scontroso, quasi sempre polemico, ma a tratti claunesco. Grazie agli amici e ai nemici, ai compagni e agli avversari, perché è alla giusta proporzione di tutti che devo la forza per portare avanti le mie battaglie. *Gracias a la vida que me ha dado tanto.* Grazie alla musica per aver dato un senso a questo mio eterno ritorno.

Un sincero ringraziamento lo rivolgo al Maestro Nicola Bernardini, una figura davvero competente che nel mio piccolo stimo davvero, insegnante capace di mettere a dura prova i suoi allievi, costringendoli a pensare prima di agire. Senza il pensiero saremmo solo carne, figuriamoci artisti.

Introduzione

Il presente elaborato verte sui processi cognitivi che caratterizzano la cosiddetta percezione cross-modale. In particolare, il nucleo d'interesse è la comprensione delle corrispondenze tra le sensazioni uditive e visive, cioè in che modo i rispettivi canali sensoriali si influenzino reciprocamente dando origine a dei paralleli percettivi.

Il punto d'avvio della presente indagine risiede in una personale inclinazione dell'autore, il quale è sovente portato a tradurre il dato uditivo in un equivalente visivo, dando vita ad un'intima dialettica fatta di suoni e immagini, una sorta di processo di autopoesi sinestesica. Da quest'attitudine, che per alcuni può apparire a dir poco triviale, si dipanano due principali percorsi paralleli che l'autore da tempo percorre con entusiasmo. Il primo è legato alla sfera della composizione musicale elettronica, dove, in modo più o meno programmatico, la produzione del materiale sonoro è frequentemente guidata da un'analogia con il mondo delle forme e della luce, così come dalla qualità fisica e sensibile della materia. A questo input iniziale segue una successiva fase di sviluppo germinativo, che può essere egualmente guidato da una data programmaticità narrativa e da una dialettica del materiale, che portano alla fioritura di texture, pattern temporali e materiale tematico. Il secondo percorso d'interesse dell'autore è, in maniera implicita, il campo delle applicazioni audio-visive, indifferentemente dalla loro natura cinematografica o installativa, dove non di rado è necessaria una cooperazione tra artisti e artisti, artisti e sound designer, sound designer e tecnici, e via dicendo, e le scelte, com'è auspicabile, devono essere guidate da una parte dal buon gusto e dall'altra da solide conoscenze.

Ecco dunque dove ha origine questa ricerca, la quale si pone l'obbiettivo di portare a emersione dei principi e dei modelli generali che guidano l'interazione percettiva fra i tratti fondamentali delle modalità uditiva e visiva. Per ottenere tali risultati ci si è preposti di produrre una revisione narrativa della letteratura scientifica che fin qui si è occupata di studi specifici, grazie alla quale giungere ad una sintesi aggiornata dell'argomento in oggetto. Una consistente parte degli studi sulla percezione cross-modale proviene dagli ambiti della psicologia cognitiva e delle neuroscienze, anche se un ristretto numero di pubblicazioni sensibili al tema è stato rintracciato nel corpus di pubblicazioni scientifiche musicali.

Considerato l'elevato numero di articoli rinvenuti, è stato ritenuto indispensabile organizzare tutto il materiale nel corso di due principali capitoli, ognuno dei quali è strutturato secondo quanto segue.

Il primo capitolo si presenta suddiviso per aree tematiche: apre con uno scorcio sulla sinestesia, tema proposto e riproposto innumerevoli volte in ambito artistico e non solo, facendo chiarezza su quali siano le peculiarità di tale fenomeno percettivo e di come possa essere messo a confronto con gli effetti meno noti (ma al contempo condivisi da tutti) della percezione cross-modale; seguirà dunque un approfondimento proprio sulla percezione cross-modale, su quali siano le sue funzioni e le sue proprietà di base. Sempre nel primo capitolo verranno affrontate brevemente le metodologie scientifiche sperimentali fin qui adottate nello studio e nella comprensione di tali processi cognitivi. Verrà poi articolata una tassonomia delle principali corrispondenze tra *features* uditive e visive: la scelta di tali corrispondenze cross-modali è stata ponderata su una verifica degli studi di provata efficacia e, dunque, le classi proposte rispecchiano quelle verificate da dati certi e inconfutabili; non si esclude che ve ne possano essere delle altre non ancora accertate, ma, allo stato dell'arte, quelle riportate nel presente elaborato rappresentano le più approfondite dalle ricerche scientifiche. Al termine del primo capitolo vengono prima riportati alcuni casi particolari che dimostrano come sia ingannabile il sistema percettivo, e infine una sezione dove verrà approfondito il rapporto tra suono e colore che, come si vedrà, non è così lineare e cristallino quanto le classi di corrispondenze cross-modali elencate in precedenza.

Nel secondo capitolo verrà proposta una bibliografia annotata delle ricerche sperimentali affrontate negli ultimi tre anni (2016-2018). A differenza del primo capitolo, questo è sviluppato secondo un criterio cronologico, con gli articoli divisi anno per anno. Per ogni articolo verrà brevemente esposto il metodo sperimentale adottato, seguito da un commento dei risultati ottenuti più salienti.

Nella speranza che una tale ricerca possa risultare trasversalmente utile su più contesti ed a più figure – siano esse compositori, sound designer, registi, artisti o appassionati di questa tematica – si augura una buona lettura.

*«La musica non si limita al mondo dei suoni.
C'è musica anche in ciò che vediamo.»*

Oskar Fischinger, 1951

Capitolo 1

Percezione cross-modale: *prior art*

1.1 Il mito sinestesico

Come un *refrain*, innumerevoli chiavi di lettura artistiche e filosofiche sono state date a un tema coinvolgente quale il rapporto tra suono, forma e colore. Il loro accostamento è da sempre stato investigato con la convinzione che un legame, al di là delle apparenze, non sia così chimerico, ma, al contrario, esso sia spiegabile attraverso dati certi.

L'immortalità della musica, “ambiguità fatta a sistema”, ha da sempre attinto anche dal mondo sensibile, quale serbatoio di forme e colori, un'infinità di suggestioni; in molte occasioni nel corso dei secoli, illustri compositori hanno scritto musica elaborando una sintassi fondata sulla co-presenza di suoni e immagini, omaggiando così l'universo dei colori e delle forme. D'altro canto si pensi al lessico e alla notazione musicale, dove vi è una forte relazione con termini utilizzati nelle arti visive, così come con quelli utilizzati per descrivere il mondo delle forme nello spazio:

alto, basso, ascendente, descendente (riferiti all'altezza); *orizzontale, posizione, intervallo, inversione* (in riferimento alla melodia); *verticale, aperto, chiuso, pesante, esile* (in riferimento all'armonia); *contrario e obliquo* (in riferimento al *contrappunto*, termine anch'esso di natura visiva). Anche la dinamica musicale conserva tracce della sua origine visiva: per esempio, linee divergenti indicano un *crescendo* o una crescita d'intensità, e linee convergenti indicano un *diminuendo*, il suo opposto.

[115, p.174]

Per certi versi è come se si fosse portati a “vedere” la musica in associazione ai movimenti di un corpo sonante nello spazio. Alle tre coordinate spaziali possono essere assegnate altrettante qualità musicali: la trama sonora (profondità), la melodia (piano verticale) e la struttura formale (piano orizzontale).

Si ricordi poi come non sia infrequente usare termini quali caldo, freddo, morbido, duro, aspro, acido, dolce, tutti in relazione al suono e alla musica; questo perché la natura effimera e astratta della musica richiede di essere personificata [45], e questa traduzione si realizza attraverso associazioni mentali rispetto al mondo fenomenico.

I legami tra il suono e la totalità dei sensi non sono così fragili, soprattutto in riferimento alla vista, come è stato ampiamente dimostrato anche dalla comunità

scientifica. Da inizio Novecento infatti, la psicologia cognitiva e le neuroscienze hanno portato avanti con determinazione e con risultati non trascurabili una lunga serie di indagini per comprendere i comportamenti di tali traslazioni sensoriali tra più dimensioni percettive. Questa proprietà, definita *percezione cross-modale* e comune a tutti, è facilmente confondibile con la ben più nota sinestesia.

In un articolo dello psicologo Lawrence Marks – direttore del John B. Pierce Laboratory e professore emerito alla Yale University – viene stilato un elenco molto completo e dettagliato di tutti gli scritti fino ad allora pubblicati (1975) sul tema della sinestesia [78]. Nell'elenco si annoverano nomi importanti quali Pitagora, Aristotele, Locke, Leibniz, Newton, Darwin, E.T.A. Hoffmann, Sachs, Baudelaire, Rimbaud, Scriabin e Kandinsky, giusto per citarne alcuni.

Già Aristotele, nel suo *De sensu et sensibilibus* (440 a.C.), proponeva un parallelismo tra l'armonia del suono e l'armonia del colore [77]. Molto più in là nei secoli, nel 1725 il padre gesuita Louis Bertrand Castel progettò il primo strumento musicale “multimediale”, un clavicembalo a colori [24], compiutamente realizzato dieci anni più tardi¹. Nel 1915, in una Carnegie Hall equipaggiata con un organo a colori, fu eseguito il poema sinfonico Op.60 di Aleksandr Skrjabin - *Prometeo: il Poema del Fuoco* (1910) - così com'era stato pensato dall'autore; in questa occasione infatti, su uno schermo posto alle spalle dell'orchestra vennero proiettati dei fasci di luce colorata prodotti dallo stesso organo, nel quale ad ogni nota era attribuito un colore (Do-rosso, Re-giallo, Mi e Fa#-blu, La-verde, Sib-grigio). Wassily Kandinsky (1866-1944), incline a una visione spirituale dell'arte, attribuiva ad ogni forma un proprio “suono interiore” e avanzava dei parallelismi tra i colori e il suono:

La tendenza del giallo ai toni chiari può raggiungere un'intensità insopportabile per lo sguardo e per l'anima. Un giallo così intenso è come il suono sempre più acuto di una tromba o quello sempre più assordante di una fanfara. [...] Da un punto di vista musicale l'azzurro assomiglia a un flauto, il blu a un violoncello o, quando diventa molto scuro, al suono meraviglioso del contrabbasso; nella sua dimensione più scura e solenne ha il suono profondo di un organo. [...] Il blu difficilmente diventa acuto e non può sollevarsi a grandi altezze. [...] Il verde assoluto è il colore più calmo che ci sia: non si muove, non esprime gioia, tristezza, passione, non desidera nulla, non chiede nulla. [...] Da un punto di vista musicale esprimerei il verde assoluto con i toni calmi, ampi, semigravi del violino.

1 Castel, Louis-Bertrand. *Nouvelles expériences d'optique et d'acoustique*, «Mémoires pour l'histoire des sciences et des beaux arts», pp. 1444-1482; 1619-1666; 1807-1839; 2018-2053; 2335-2372; 2642-2768.

[...] il bianco, che spesso è considerato un non-colore [...], è un mondo così alto rispetto a noi, che non ne avvertiamo il suono. [...] Il grigio è silenzioso e immobile. La sua immobilità, però, è diversa dalla quiete del verde, che è circondata e prodotta da colori attivi. [66, p.62 e seguenti]

Chiaramente la terminologia, gli strumenti ed i metodi utilizzati in passato per approfondire i ponti tra l'udito e la vista non sono paragonabili a quelli di cui si serve la scienza contemporanea, ma è comunque interessante notare come questa tematica abbia da sempre appassionato artisti, scrittori, pensatori e in generale i grandi protagonisti della cultura occidentale. Dopo essere stata per secoli fonte di inspiegabile stupore e di cogitazioni multiformi, verso la fine dell'Ottocento il tema della sinestesia aveva raggiunto una notorietà tale che venne preso sul serio dalla psicologia, anche se il primo studio ufficiale risale al 1812, anno in cui il medico Georg Sachs pubblicò un pionieristico reportage sull'audizione colorata (*audition colorée*) [24]. Nel 1890, la sessione di apertura dell'*International Congress of Psychological Psychology* di Parigi fu dedicata a un dibattito tenuto da un comitato di sette prominenti psicologi il quale propose una tassonomia delle varie esperienze sinestetiche e una definizione di audizione colorata [78]. Gradualmente, nel corso del Novecento l'interpretazione patologica della sinestesia come forma di anormalità fu abbandonata in favore di un'interpretazione fisiologica e psicologica che vede la sinestesia come una normale funzione cognitiva, il prodotto di un'associazione mentale di dati sensoriali [24]. Tuttavia, al fine di scongiurare equivoci terminologici, è necessario avanzare una distinzione tra sinestesia "vera" e percezione cross-modale.

La sinestesia indica un fenomeno psichico per cui la stimolazione di una determinata modalità sensoriale dà origine a un'esperienza in una seconda modalità percettiva [59]. Quindi un singolo stimolo (*inducer*, evento induttore iniziale) produce una duplice sensazione, investendo così anche altri canali sensoriali non direttamente sollecitati (*concurrent*, evento concorrente). La parola *sinestesia* deriva infatti dall'unione dei termini greci *syn* (assieme) e *aisthesis* (percepire) [24]. La percentuale della popolazione mondiale in cui si riscontrano un comportamento sinestetico viene stimata tra lo 0,5 e il 5% [27], e i casi più frequenti sono legati alla sollecitazione della modalità uditiva, da cui scaturisce la produzione di immagini interiori (la cosiddetta *audizione colorata* o *chromesthesia*). Gli effetti sinestetici possono essere congeniti, ma possono anche comparire in seguito a danni cerebrali [42].

Nella *chromesthesia* uno stimolo uditivo (*inducer*) produce dei fotismi (*concurrent*), cioè delle immagini colorate [78]. Sempre partendo dagli stimoli

uditivi, non sono poi così rari (anche se proporzionalmente in numero minore rispetto alla *chromaesthesia*) i casi in cui vengano attivati anche altri organi sensoriali rispetto alla vista, come sostiene Oliver Sacks:

Mentre ascoltano la musica, alcune persone – un numero sorprendentemente alto - «vedono» colori o provano varie sensazioni «gustative», «olfattive» o «tattili»: tale sinestesia, d'altra parte, può essere considerata tanto un sintomo, quanto un dono. [113, p.17]

Per poter parlare di sinestesia vera e propria devono essere verificati una serie di comportamenti [16], quali:

- attivazione involontaria e automatica del processo cognitivo;
- produzione di precise sensazioni soggettive invariabili nel tempo;
- produzione di precise sensazioni che vanno a sommarsi alla normale percezione, non a sostituirla e nemmeno a contraffarla con effetti di mascheramento.

Recenti studi hanno però dimostrato come i processi associativi involontari tra più modalità percettive tipici di un ristretto numero di sinesteti, siano molto simili a quelli originati nella normale percezione cross-modale [143]. Appositi test, sfruttando le tecniche di risonanza magnetica (fMRI, *Functional Magnetic Resonance Imaging*), mostrano una maggiore attivazione a livello neurale nei soggetti sinestetici rispetto ai non-sinesteti quando entrambi vengono sottoposti a stimoli esterni [95], ma le aree corticali attivate sono le medesime. Le funzioni percettive cross-modali sono dunque comuni a tutti gli esseri umani, sebbene le sensazioni cross-modali generiche siano più blande e meno precise paragonate a quelle sinestetiche.

A distanza di oltre un secolo dalle riflessioni astratte e di stampo spiritualista avanzate da Kandinsky, oggi, grazie a metodi scientifici accurati, è possibile dimostrare la fondatezza di fenomeni associativi tra suono, forma e colore, classificabili come processi di percezione cross-modale.

1.2 Percezione cross-modale

La capacità di integrare le informazioni provenienti dall'ambiente è una caratteristica fondamentale del cervello umano [129]. Tale organo combina continuamente la complessa varietà di stimoli percepiti dai diversi canali sensoriali (vista, udito, olfatto, gusto e tatto), con la conseguente attivazione di risposte appropriate da parte dell'organismo [35]. Tale ridondanza di informazioni e il loro filtraggio selettivo sta alla base dell'*integrazione cross-modale* (o *convergenza cross-modale*), intesa come l'unificazione dei diversi dati sensoriali in un'unica informazione multidimensionale, la quale viene poi associata ad un singolo oggetto o evento fisico esterno. Lawrence Marks, in scia con i precetti ecologici di James Gibson [53], parla di *unità dei sensi* (*The Unity of the Senses*), per indicare la sinergia dei vari organi di senso, i quali condividerebbero delle qualità fenomenologiche comuni [79]. All'integrazione cross-modale viene attribuita anche una *funzione adattiva* [2], orientata cioè al miglioramento della qualità nella rilevazione di oggetti ed eventi circostanti in termini di velocità e precisione.

Le funzioni qualitative della convergenza cross-modale sono dunque intrinseche all'attività percettiva. Ma vi è un ulteriore aspetto, centrale in questo scritto, che caratterizza tali rapporti intra-sensoriali di base: numerosi studi stanno dimostrando in maniera sempre più consistente secondo quali criteri psico-percettivi l'essere umano sia in grado di elaborare delle *corrispondenze cross-modali* tra più stimoli appartenenti a diverse modalità sensoriali [127]. Ad esempio, si è generalmente portati ad associare suoni acuti a oggetti piccoli, luminosi e spazialmente elevati. Queste associazioni sono elaborate a livello subconscio, e possono essere considerate delle *metafore percettive*, poiché descrivono delle semplici associazioni prodotte dell'immaginazione (anche se a volte molto vivide), ma non costituiscono delle co-sensazioni in senso stretto [45] come avviene invece nel caso dei processi sinestetici. Tuttavia va tenuto presente che queste associazioni non sono frutto di scelte casuali e arbitrarie, ma sottostanno a certe leggi governate da processi cognitivi comuni a tutti, fatto che ne permette lo studio e la misurazione attraverso appositi test percettivi; tali esperimenti hanno lo scopo di spiegare e classificare secondo precisi criteri le proprietà e il funzionamento dei processi cognitivi cross-modali.

Sebbene ogni persona nell'arco della propria vita maturi una propria esperienza sensoriale in rapporto al mondo², le corrispondenze cross-modali sono processi congeniti e sono state verificate già nei primi anni di vita [82]. È stato infatti dimostrato come già nella prima infanzia siano riscontrabili effetti di corrispondenze cross-modali tra varie proprietà degli stimoli uditivi e visivi: vengono infatti prodotte delle associazioni tra l'intensità sonora e la luminosità della luce [73] o le dimensioni delle forme [125], tra l'altezza dei suoni e le caratteristiche geometriche delle forme (quali la loro rotondità/spigolosità, l'elevazione spaziale e la magnitudine delle loro dimensioni [94][141]), così come anche tra l'altezza sonora e la luminosità della luce [94].

La vista è considerata la modalità percettiva dominante [11] (chiaramente per i soggetti che non presentano deprivazioni visive e sensoriali), ma gli altri sensi, in primo luogo l'udito, interagiscono con essa per generare una rappresentazione cognitiva coerente degli eventi fisici circostanti [27]. Nel reciproco scambio di informazioni tra i sensi, i punti di forza di una modalità riescono a compensare i limiti delle altre [96].

Nella continua percezione di eventi acustici, l'uomo fa incessantemente uso di corrispondenze cross-modali, e in rapporto alla musica non vi è eccezione: la sua fruizione sprigiona un via vai di fantasmagorie d'ogni tipo, anche se, in certa misura, i rimandi possono essere connessi all'esperienza del mondo fenomenico, il quale ritorna sistematicamente con le sue immagini a corredo di ciò che viene udito. È lecito dunque affermare che l'atto di guardare o ascoltare possa essere inteso come un'attività creativa di costruzione di oggetti [58]; tale idea viene condivisa, tra gli altri, anche dal compositore Denis Smalley:

Our sense of texture is learned through vision and touch as well as sound; our experience of the physical act of sound making involves both touch and proprioception; spectral motion, and the movement and distribution of sounds in space relate to our own experience of physical motion and cultural and natural environments. [...] Transmodal linking occurs automatically when the sonic materials seem to evoke what we imagine to be the experience of the world outside the music [...]³. [124, p.39]

-
- 2 La *plasticità sinaptica* è la capacità del sistema nervoso di modificare la sua organizzazione a causa di molti eventi, che vanno dal normale sviluppo e la relativa maturazione dell'organismo, dall'acquisizione di nuove abilità (“apprendimento”) in organismi immaturi e maturi, ma anche conseguentemente al danneggiamento dello stesso sistema nervoso o alla deprivazione sensoriale [5].
- 3 La nostra sensazione delle texture è appresa attraverso la vista e il tatto, così come attraverso il suono; la nostra esperienza dell'atto fisico della creazione del suono coinvolge sia

Il presente scritto verte sulle corrispondenze tra due canali percettivi, quello uditivo e quello visivo; più precisamente, si cerca di comprendere come la percezione di eventi sonori sia in grado di influire sulla percezione di stimoli visivi, e altresì come le qualità tipiche dei tratti sonori riescano a trovare delle corrispondenze e dei paralleli nelle proprietà appartenenti alla modalità visiva. Va comunque tenuto presente che le ricerche hanno evidenziato l'esistenza di molte altre corrispondenze tra coppie sensoriali, quali udito e tatto [149], udito e olfatto [7][117], udito e gusto [6][21][114] [122], vista e tatto [89], vista e olfatto [67], vista e gusto [44].

Dagli studi sulla percezione cross-modale si evincono due aspetti principali dei processi cognitivi; la mente da una parte è in grado di operare una cosiddetta *congruenza semantica* (*semantic congruency*) nei casi in cui la percezione di una coppia di stimoli (ad es. uditivi e visivi) viene qualificata sulla base di una reciproca congruenza (o incongruenza) in termini di significato semantico, mentre dall'altra è in grado di elaborare una cosiddetta *congruenza sinestetica* (*synaesthetic congruency*) nei casi in cui una caratteristica di uno stimolo percepito (ad es. l'altezza o l'intensità di un suono) viene associata ad un attributo di un secondo stimolo "virtuale", cioè non percepito fisicamente come il primo, bensì solamente immaginato [127].

Con *unity effect* [126][139] si fa riferimento ai casi in cui due o più input sensoriali vengono percepiti come altamente coerenti; in tale circostanza vi è la propensione ad assumere tali stimoli come aventi un'origine spazio-temporale comune, e di conseguenza tutti gli input vengono associati ad un singolo oggetto o evento fisico. Il soggetto percipiente avrà dunque la sensazione di distinguere un singolo evento multimodale rispetto a molteplici eventi unimodali basandosi sulla congruenza delle informazioni disponibili in ciascuna modalità sensoriale. Sarà dunque a discrezione del soggetto, sulla base della congruenza semantica dei vari input, valutare quali di essi debbano essere uniti assieme e associati ad un singolo evento (*crossmodal binding problem* [127]). Ciò dipende anche dal tipo di caratteristiche dell'evento/oggetto rilevato, caratteristiche che possono essere di due tipologie: *amodali*, qualora esse possono essere percepite da più canali sensoriali (ad es. la forma di un solido può essere vista, ma anche toccata), o *modali*, se peculiari di un singolo canale percettivo (ad es. l'altezza di un suono o il colore di una luce).

il tatto che la propriocezione; il movimento spettrale, il movimento e la distribuzione dei suoni nello spazio si connettono alla nostra esperienza del movimento fisico e degli ambienti culturali e naturali. [...] Il collegamento trans-modale si verifica automaticamente quando i materiali sonori sembrano evocare ciò che immaginiamo essere l'esperienza del mondo al di fuori della musica.

Le corrispondenze cross-modali possono verificarsi nei modi più vari. Il caso più elementare avviene in presenza di stimoli amodali che, come specificato poco sopra, coinvolgono più organi di senso; caratteristiche apparentemente incongruenti (anche modali) e percepite da più recettori sensoriali possono essere comunque correlate, così come nel caso limite della congruenza sinestetica, ove un singolo stimolo può essere associato in modo astratto ad un altro oggetto/evento secondo un principio di affinità tra le rispettive caratteristiche modali o amodali, come avviene ad esempio nell'abbinamento di suoni acuti a oggetti piccoli e luminosi.

Nella letteratura scientifica abbondano i termini con i quali vengono definite le corrispondenze cross-modali: corrispondenze sinestetiche, associazioni sinestetiche, equivalenze cross-modal, similitudini cross-modal, ecc. Le espressioni legate alla sinestesia risultano tuttavia fuorvianti poiché, così com'è stato appurato nel precedente paragrafo, la sinestesia è un fenomeno ben preciso che, sebbene per certi versi possa essere paragonato alla percezione cross-modale, presenta comunque delle sfumature diverse.

Il termine *corrispondenza cross-modale* verrà qui adottato in riferimento agli effetti associativi fra i tratti di più stimoli appartenenti a diverse modalità sensoriali; tali corrispondenze implicano dunque che le qualità percettive di uno stimolo, su una data dimensione, risultino compatibili o equivalenti alle qualità percettive di uno stimolo di diversa natura sulla dimensione corrispondente, sia esso reale o virtuale (nel senso di immaginato). Il presupposto è che tali corrispondenze cross-modali siano condivise da un gran numero di persone, e alcune potrebbero, di fatto, essere universali [127].

Il legame tra la percezione cross-modale e la sinestesia, al di là degli equivoci terminologici, risulta essere comunque importante per due motivi: da un punto di vista teorico, poiché i comportamenti mostrati dai sinesteti e dai non-sinesteti sono molto simili [3][88]⁴, e da un punto di vista pratico poiché, proprio per la similitudine tra sinesteti e non, nello studio delle corrispondenze cross-modal una consistente parte dei test percettivi mette a confronto i risultati ottenuti rispettivamente da gruppi

⁴ Gail Martino e Lawrence E. Marks hanno evidenziato come le corrispondenze cross-modal mostrino molti tratti in comune rispetto alla vera sinestesia, suggerendo la possibilità che entrambi i processi percettivi condividano gli stessi meccanismi neurali (pensiero condiviso anche in [143]); per questo motivo essi proposero una distinzione tra *sinestesia forte* (*Strong Synesthesia*), caratterizzata da stimoli concorrenti vividi, e *sinestesia debole* (*Weak Synesthesia*), caratterizzata da somiglianze inter-sensoriali meno intense e dirette, solitamente espresse tramite il linguaggio.

di soggetti sinestetici e da gruppi di controllo (persone cioè che non presentano comportamenti sinestesici).

Jamie Ward e Jason B. Mattingley, attraverso tre diversi test sull'audizione colorata che mettevano a confronto un gruppo di sinesteti e un gruppo di controllo, danno prova della somiglianza tra la percezione cross-modale e la sinestesia [143]. Sottoponendo entrambi i gruppi a medesimi stimoli uditivi e chiedendo ai soggetti di abbinare le tonalità cromatiche ritenute congrue al dato sonoro, le differenze che emergono dai test sono più che altro legate al grado di coerenza con il quale i sinesteti selezionano le feature visive (sempre immutate), con reazioni più immediate rispetto a quelle empiriche dei non sinesteti. Tuttavia, poiché sinesteti e non fanno uso dei medesimi processi cognitivi cross-modali di base, l'associazione tra il dato sonoro e il colore risulta molto simile.

Un altro test con sinesteti e non a confronto, è stato effettuato da Pineas de Thornley Head [61] sempre per verificare il grado di corrispondenza tra suono e colore. Anche in questo caso il risultato più saliente sta ad indicare una cosa abbastanza ovvia, e cioè che per i sinesteti, nonostante vengano sottoposti a stimoli ripetuti a distanza di tempo, i fotismi rimangono invariati; per i non-sinesteti invece è più difficile riportare gli stessi risultati invariati nel tempo, in quanto le immagini suscite sono più vaghe rispetto ai fotismi dei sinesteti. Nei non-sinesteti subentrano infatti anche altri fattori interpretativi quali l'immaginazione, la memoria e le metafore. Lawrence Marks definisce *metafore sinestetiche* [80][83] quelle espressioni in cui parole o frasi che tipicamente descrivono esperienze proprie di un senso, vengono usate per trasferire i loro significati ad un'altra modalità sensoriale (ad es. “suono ruvido”, “suono acido”, “colore caldo”, “colore squillante”).

Tra gli altri test che vedono contrapposti gruppi di sinesteti e gruppi di controllo, si segnala quello riportato da Chiou [27]; in questo caso sono stati sottoposti entrambi i gruppi a stimoli uditivi di varia natura con l'obiettivo di verificare il grado di ricorsività e di affinità nelle corrispondenze tra suono, colore e posizione spaziale. Dal test si evince in primo luogo una similarità nei processi associativi dei due gruppi; questo combacia con l'esito del test di Ward e Mattingley [143] e avalla ulteriormente la tesi secondo cui i processi percettivi di base siano condivisi da tutti. Inoltre si evince chiaramente una correlazione sistematica tra l'altezza del suono e i tratti geometrici degli oggetti e la loro collocazione spaziale: più i suoni sono acuti, più le forme evocate sono di colore chiaro, di dimensioni ridotte ed elevate nello spazio. Al contrario, i suoni gravi evocano forme di colore più scuro, di dimensioni più grandi e collocate in posizioni più basse.

Queste associazioni hanno indubbiamente origine anche dall'esperienza che normalmente si acquisisce in rapporto al mondo fisico [45], dove l'altezza del suono prodotto dagli oggetti è legato innanzitutto alle loro dimensioni: corpi grandi producono generalmente suoni gravi, viceversa quelli di piccole dimensioni danno origine a suoni acuti.

Un'altra correlazione molto diffusa è quella tra la luminosità del colore e l'altezza del suono: un tono basso appare più cupo di uno alto. Nel mondo fisico l'opacità di un corpo spesso è legata alla sua densità e di conseguenza al suo peso specifico; da ciò si deduce che gli oggetti più scuri vengono associati a corpi più pesanti, che tendono così a cadere verso il basso, interpretazione che spiegherebbe la relazione tra luminosità e altezza sonora.

Il peso di un oggetto ha poi una conseguenza sulla forza d'impatto prodotta nella collisione con un altro corpo rigido; questo spiega perché si è portati a relazionare immagini di grandi dimensioni a suoni ad alta intensità, laddove quelli deboli evocano invece immagini minute.

Nei test percettivi i partecipanti vengono sottoposti a stimoli di varia natura ed esprimono le relative associazioni sonoro-visuali attraverso comunicazioni verbali (congruenza sinestetica) oppure segnalando uno stimolo ritenuto più coerente scelto tra un set di stimoli dato (congruenza semantica). Se chiamati a descrivere le sensazioni visive prodotte dagli stimoli acustici, per i soggetti dei gruppi di controllo è difficile manifestare delle rappresentazioni dettagliate, con colori, forme, dimensioni e posizioni spaziali ben precise e invariabili nel tempo; essi si limitano a delle descrizioni approssimative e comparative tra estremi morfologici. Questo è spiegabile dal fatto che il mondo fenomenico non si basa sulla magnitudine complessiva dei suoi eventi (ad esempio luce e suono), ma sulla differenza o il rapporto tra gli opposti (chiaro-scuro, duro-morbido, alto-basso, grande-piccolo, ecc.). In termini generali, questo contrasto permette di suddividere e classificare - attraverso una codifica semantica fatta di corrispondenze relazionali [82] - l'universo percettivo negli oggetti e negli eventi con i quali si interagisce. Inoltre, tale contrasto consente di collocare oggetti ed eventi in categorie di equivalenza, necessarie per dare un senso al mondo circostante [58].

1.3 Metodologie sperimentali per lo studio della percezione cross-modale

In questo paragrafo ci si concentra brevemente sui metodi impiegati nello studio della percezione cross-modale, allo scopo di fare un pò di chiarezza su tali modelli sperimentali, facilitando così la comprensione dei paragrafi a seguire ed eventuali approfondimenti del lettore.

Il mondo fisico è costituito di materia, energia e fenomeni fisici, ed è strutturato secondo leggi proprie; l'essere umano che lo abita non è un semplice "ricevitore passivo" ma, al contrario, si impegna attivamente ad acquisire informazioni, anche quelle non immediatamente disponibili, selezionando e analizzando la struttura di ciò che lo circonda nell'ambiente. In altre parole, si comporta in modo critico. Tutta la recente letteratura e gli studi sulla percezione cross-modale rientrano nel campo del cognitivismo e delle neuroscienze, e possono essere fatti risalire all'opera iniziativa dello psicologo americano Wendell R. Garner⁵ (1921 – 2008) [84], uno dei padri della psicologia cognitiva che, grazie al suo programma di ricerca iniziato negli anni sessanta, contribuì a dare vita alla scienza cognitiva⁶. La psicologia cognitiva è la branca della psicologia che studia i processi mentali attraverso i quali le informazioni vengono acquisite, elaborate, memorizzate e recuperate dal sistema cognitivo. Per sistema cognitivo generalmente ci si riferisce a tutti i meccanismi mentali legati alla conoscenza (percezione, immaginazione, memoria, ragionamento) che guidano il comportamento dell'uomo.

Brevemente, il programma di ricerca di Garner ha inizialmente cercato di chiarire le caratteristiche ed i principi base dell'*attenzione selettiva* dell'uomo nei confronti degli stimoli esterni, nel tentativo di comprendere in che misura le persone prestino attenzione a particolari attributi di un medesimo stimolo mentre per contro ne ignorino gli altri. Gli studi presero le mosse chiarendo i comportamenti selettivi tipici dei singoli canali percettivi; ad esempio, nella modalità uditiva quale sia l'interazione

5 Per un approfondimento sugli studi di Garner si veda il suo importante volume [48]: Garner, Wendell R. *The processing of information and structure*, L. Erlbaum Ass., New York, 1974.

6 La scienza cognitiva ha come obiettivo lo studio dei processi cognitivi, umani e artificiali, e si compone di una serie di discipline quali: intelligenza artificiale, psicologia cognitiva, linguistica, psicolinguistica, filosofia della mente e del linguaggio, neuroscienze e antropologia.

tra l'altezza di un suono e la sua intensità. Successivamente, i test che iniziarono a studiare i fenomeni della percezione cross-modale, in un certo senso aumentarono il grado di complessità, poiché si posero l'obiettivo di capire quali siano e come avvengano le interazioni tra le rispettive caratteristiche di due stimoli appartenenti a canali sensoriali diversi, come ad esempio quello visivo e quello uditivo.

Per tale scopo sono stati sviluppati due principali modelli di test percettivi, chiamati rispettivamente *baseline task* (compiti di base) e *selective attention task* (compiti di attenzione selettiva). In ogni test vengono studiati i comportamenti dell'attenzione selettiva rivolti nei confronti di una specifica caratteristica (*feature*) appartenente ad una determinata categoria di stimoli; tale *feature* viene dunque considerata il tratto rilevante da studiare (R) (ad es. il pitch nel caso di stimoli acustici). La caratteristica rilevante viene solitamente messa a confronto con un'altra *feature* considerata irrilevante (I), la quale può appartenere allo stesso genere di stimolo (ad es. la loudness) oppure ad uno stimolo appartenente ad un'altra modalità percettiva (ad es. la luminosità del colore), così come avviene negli studi sulla percezione cross-modale.

Nei *baseline task* i soggetti partecipanti sono sottoposti a una serie di stimoli che variano nella dimensione rilevante (R), mentre rimangono costanti nella dimensione irrilevante (I). Si prenda come esempio il caso in cui si cerchi di comprendere l'influenza del pitch sulla percezione della luminosità: la dimensione rilevante sarà dunque il pitch, presentata con due o più segnali acustici a varie altezze (nel caso più semplice, un suono grave e uno acuto), mentre quella irrilevante sarà la luminosità, per cui ad ogni stimolo acustico verranno mostrate immagini con lo stesso identico valore di luminosità. Ai partecipanti viene dunque chiesto di identificare quale stimolo tra quelli nella dimensione rilevante consideri più in accordo rispetto agli stimoli nella modalità percettiva irrilevante; per questo i *baseline task* sono anche definiti test identificativi.

Nel caso dei *selective attention task* invece, entrambe le dimensioni (rilevante e irrilevante) vengono variate, ed ai soggetti viene chiesto di classificare come corretta o meno l'associazione tra le due dimensioni, motivo per il quale questi test vengono definiti test classificativi. Talvolta questi test vengono anche chiamati *filtering task*, poiché si presume che per classificare correttamente le coppie di stimoli i soggetti debbano effettuare un'operazione di filtraggio degli stimoli lungo la dimensione irrilevante.

Un ulteriore dato importante può essere estrapolato dal monitoraggio dei tempi di reazione dei partecipanti agli stimoli: nei cosiddetti *speed test* viene infatti

misurato il tempo con il quale i soggetti riescono a completare le singole prove a cui sono sottoposti (RT, *reaction time*). Anche se non tutti i test sono pensati per valutare i tempi di reazione, il modello degli speed test risulta quello più adottato negli ultimi trent'anni [104]. I tempi di reazione dei partecipanti sono spesso più veloci per determinate combinazioni di stimoli (rilevanti e irrilevanti) e più lenti per gli altri. Sulla base di questa differenza nei tempi di reazione, le combinazioni di stimoli che portano a RT più bassi sono considerate congruenti, mentre quelle che portano a RT più lunghi sono considerate incongruenti. Gli speed test tendono dunque ad evidenziare il grado di sensibilità delle corrispondenze cross-modali, mentre i test che non prevedono la misura dei tempi di reazione tendono a risaltare maggiormente il grado di precisione delle singole feature. Per i test senza verifica dei tempi di reazione, un'ulteriore variante rispetto ai *baseline task* ed ai *selective attention task* è rappresentata dai *crossmodal matching task*, test dove sono gli stessi partecipanti a settare la magnitudine della *feature* rilevante sulla base della corrispondente *feature* irrilevante di un secondo stimolo; comprensibilmente i dati rilevati in tali test risultano ancora più accurati e maggiormente rappresentativi dei processi cross-modal.

Qualora le risposte dei soggetti sottoposti ai test di percezione cross-modale risultino accurate, veloci e invariabili, si parla di *effetto congruenza (congruence effect)* [84]; ovviamente ciò si verifica quando gli stimoli appartenenti a due diverse modalità percettive vengono percepiti come corrispondenti, motivo per il quale si può parlare di *congruenza cross-modale*.

In alcuni casi i soggetti faticano a mostrare in modo chiaro degli effetti di congruenza tra due *features* di stimoli appartenenti a modalità diverse, soprattutto quando uno dei due stimoli presenta un certo grado di intrinseca complessità. Il caso più frequente si verifica in quei test dove ad uno stimolo uditorio viene contrapposto uno stimolo colorato; il colore infatti, come si vedrà meglio più avanti, si compone di per sé di più dimensioni percettive (tonalità, luminanza e saturazione), per cui risulta difficile giudicare una dimensione rilevante del suono quando per contro vengono presentati colori su una gamma cromatica più o meno vasta. I soggetti sono invece più facilitati nelle risposte qualora vengano presentati stimoli colorati variati in una sola dimensione (luminanza o saturazione). Quando gli effetti di congruenza cross-modale studiati sulla dimensione rilevante di uno stimolo vengono minati da una dimensione irrilevante di uno stimolo che allunga i tempi di risposta e ne diminuisce l'accuratezza, si ha la cosiddetta *interferenza di Garner* [107]. Riprendendo l'esempio del colore, risulta infatti più difficile discriminare in modo selettivo la luminosità di uno stimolo visivo quando lo stimolo varia in modo imprevedibile anche nella

tonalità o nella saturazione; la variazione di questi due tratti del colore produce infatti un effetto interferenza. Le dimensioni percettive che minano reciprocamente una loro corretta valutazione sono definite *integrali* [48], mentre vengono definite *separabili* tutte quelle dimensioni che possono essere giudicate in maniera selettiva e percepite in modo univoco senza interferenze da parte di ulteriori *features* dello stesso stimolo.

Per quanto concerne il rapporto tra le sensazioni uditive e quelle visive, i migliori esempi di features separabili sono quelli che verranno presi in considerazione nel successivo paragrafo dedicato alle corrispondenze cross-modali uditivo-visive.

1.4 Tassonomia delle corrispondenze crossmodali uditivo-visive

Un consistente numero di studi basati su test psico-percettivi hanno stabilito con precisione l'esistenza di corrispondenze cross-modali tra la modalità uditiva e quella visiva. Volendo stilare un elenco di tali corrispondenze è necessario stabilire una serie di binomi tra le *features* del suono e quelle visive; ad ogni caratteristica del suono, dunque, viene fatta corrispondere una certa qualità visiva.

Nella fattispecie, le correlazioni più consistenti sono:

| <i>Percezione uditiva</i> | <i>Percezione visiva</i> |
|---------------------------|---------------------------------|
| Pitch | Luminosità |
| Pitch | Dimensione delle forme |
| Pitch | Elevazione spaziale delle forme |
| Pitch | Geometria delle forme |
| Loudness | Luminosità |
| Loudness | Dimensione delle forme |
| Timbro | Geometria delle forme |

Table 1: Principali corrispondenze crossmodali tra la modalità uditiva e visiva.

Prima di passare in rassegna tali corrispondenze cross-modali tra suono e immagini, si riportano brevemente alcune definizioni terminologiche utili per una corretta comprensione dei punti a seguire. In questa sede interessano tanto le *dimensioni fisiche* quanto le dimensioni *percettive* dei fenomeni uditivi e visivi; se da una parte infatti ogni evento fisico è misurabile con apposite grandezze, ogni sua misura può variare una volta traslato dal mondo fisico a quello percettivo. Per quanto riguarda il suono, ad esempio, la misura della sua intensità (dB) quale fenomeno fisico energetico è diversa rispetto alla percezione che si ha di essa, poiché orecchio e cervello cooperano elaborando e traducendo il suono secondo leggi proprie. Dimensioni fisiche e dimensioni percettive, com'è intuibile sono strettamente correlate, ma vengono indicate con nomi diversi.

Le dimensioni fisiche e percettive del suono più rilevanti in questo scritto possono essere ricondotte a tre parametri fondamentali: *loudness*, *pitch* e *timbro*. La *loudness* (*phon*) è una variabile psicoacustica che si riferisce alla percezione soggettiva della pressione sonora dell'onda acustica, variabile anche in base alla sua frequenza, così come viene evidenziato dalle curve isofoniche del grafico sviluppato da Fletcher e Munson [41] riportato sotto.

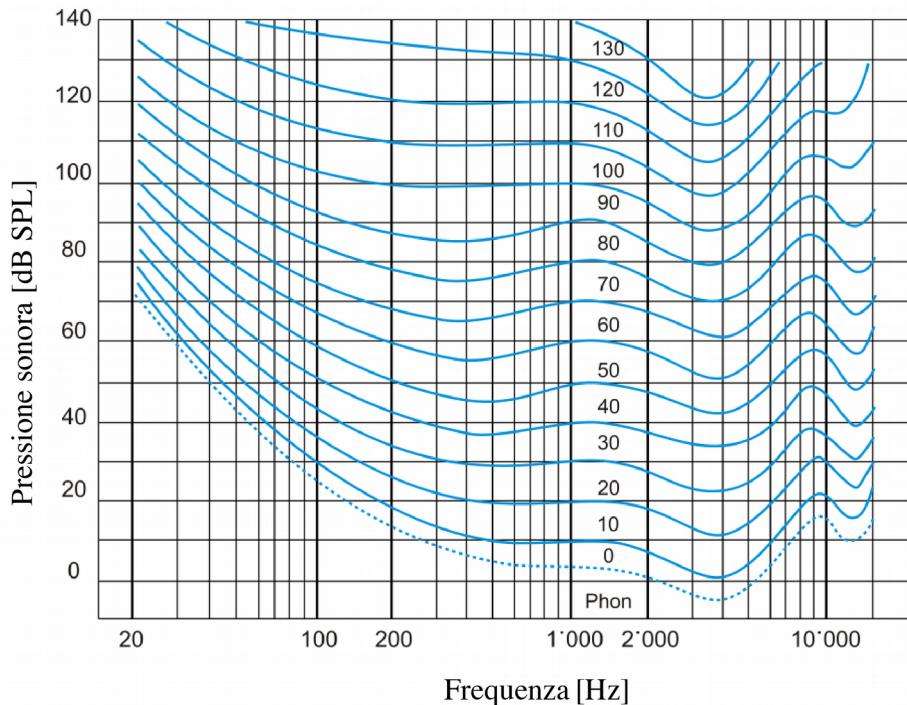


Illustration 1: Curve isofoniche di Fletcher e Munson.

Il *pitch* è legato alla sensazione di quanto è acuto o grave un suono (altezza), e dipende principalmente dalla sua frequenza fondamentale, cioè il numero di ripetizioni al secondo del suo pattern di vibrazione; tra i suoni periodici possono essere distinti i suoni puri da quelli complessi: i suoni puri sono costituiti da una sola componente frequenziale, mentre quelli complessi sono dati da una sovrapposizione di più componenti (ognuna con una sua frequenza, altezza e fase), per i quali la frequenza fondamentale esprime il massimo comun divisore nella serie di frequenze che costituiscono lo spettro sonoro. Il *timbro* è quel tratto del suono in base al quale due suoni complessi aventi stessa altezza, durata e intensità vengono giudicati diversi. Ciò che caratterizza il *timbro* è lo spettro energetico del suono e la sua evoluzione

temporale; in riferimento a quest'ultimo punto, è noto come eliminando l'attacco di un suono strumentale (ad esempio una nota di pianoforte), la riconducibilità di tale suono a un determinato strumento risulti compromessa.

Per quanto riguarda la percezione visiva, essa è normalmente in grado di integrare le informazioni relative al colore, alla forma e alla texture di oggetti e superfici [52]. Il colore, così com'è stabilito dal sistema Munsell (utilizzato come standard internazionale), è definibile come la relazione di tre coordinate dimensionali: la tonalità (*hue*), la brillanza (*lightness*) e la saturazione (*chroma*). Per *tonalità* ci si riferisce al colore “puro” percepito dall’occhio, determinato dalla lunghezza d’onda dominante all’interno dello spettro di luce visibile. La *brillanza* è la luminosità percepita di un colore tra gli estremi luminosità-buio in una scala da 0 (nero) a 10 (bianco). L’equivalenza fisico della brillanza viene chiamata *luminanza*, misurata in cd/m², cioè una grandezza fotometrica misurabile come il rapporto tra l’intensità luminosa emessa da una superficie (cd) e l’area di tale superficie (m²). Nelle pagine a seguire verrà utilizzato il termine *luminosità* per unificare entrambe le caratteristiche (quella fisica, *luminanza*, e quella percettiva, *brillanza*). La saturazione rappresenta l’intensità del colore, cioè la misura di quanto esso differisce dal grigio [102]. Un colore molto saturo ha un aspetto vivido, mentre al diminuire della sua saturazione esso si spegne e tendendo al grigio.

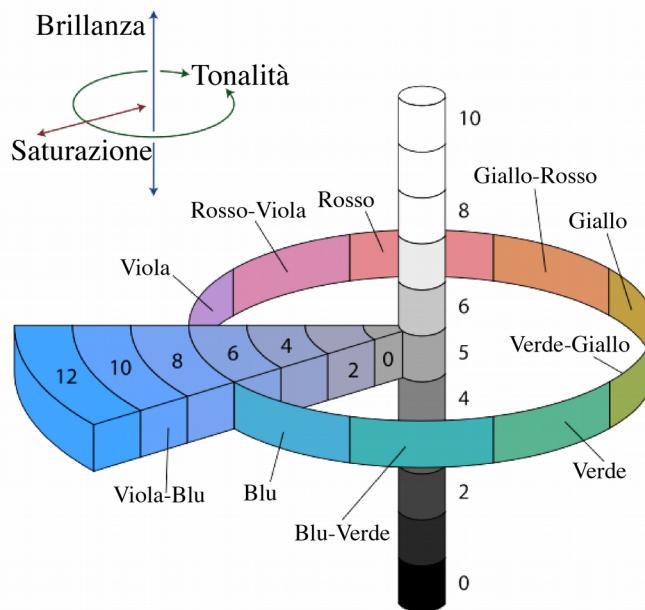


Illustration 2: Sistema Munsell dei colori.

1.4.1 Pitch – Luminosità

Questa classe di corrispondenze cross-modali è una tra le più studiate in assoluto; i test stabilito che la corrispondenza tra l'altezza sonora e la luminosità è la più saliente e coerente a livello percettivo [78][82][85]. È stato dimostrato che vi è la generale tendenza a correlare l'aumento di altezza del suono (a parità di loudness) all'aumento della brillanza (luminosità percepita) [77][79][80]. L'elaborazione cross-modale tra il pitch e la luminosità rivela così un'associazione direttamente proporzionale tra queste due *features*.

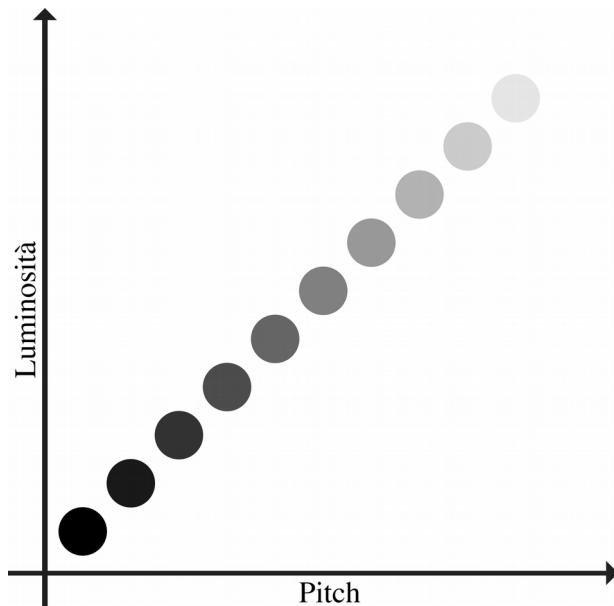


Illustration 3: Modello base per la corrispondenza cross-modale pitch-luminosità.

Da un punto di vista percettivo dunque, gli stimoli visivi più luminosi si adattano meglio alle sonorità più acute, mentre quelli più scuri sono in accordo con i suoni più gravi. Questa congruenza di base risulta più marcata quando gli stimoli visivi sono collocati su uno sfondo nero; viceversa, con uno sfondo bianco i soggetti mostrano tempi di reazione agli stimoli più lunghi [62]. Tuttavia, tale effetto di disturbo del campo visivo viene eliminato quando ai partecipanti viene data la possibilità di selezionare autonomamente il livello di luminanza ritenuto più congruente all'altezza del segnale acustico dato.

La congruenza tra altezza e luminosità è confermata anche dagli *speed test*, i quali rivelano tempi di reazione decisamente più bassi quando suoni acuti vengono presentati simultaneamente a stimoli visivi con alti valori di luminanza, oppure, in modo speculare, quando suoni gravi sono rapportati a immagini più scure; i tempi di reazione invece si allungano quando le relazioni tra stimoli uditivi e visivi sono disarmoniche rispetto al modello di base mostrato in figura 3 [81].

Qualora venga chiesto di individuare una corrispondenza tra la luminosità di uno stimolo visivo e un certo intervallo melodico, agli stimoli più luminosi vengono attribuiti gli intervalli melodici ascendenti, viceversa gli stimoli più scuri sono associati a intervalli melodici discendenti [62].

La corrispondenza cross-modale tra pitch e luminosità viene rilevata anche a livello linguistico, secondo la logica delle metafore sinestetiche; in un test di Lawrence Marks ai soggetti partecipanti veniva chiesto di classificare un set di parole date e un altro set di espressioni metaforiche che associano suono e immagini secondo la loro intrinseca luminosità [81]. *Sunlight* era la parola considerata più luminosa, mentre *black* risultò quella più scura. L'espressione metaforica considerata più luminosa era *sunlight roars* (la luce del sole ruggisce), seguita da *sound of sunlight* (il suono della luce del sole); *dark thunder* (tuono scuro) fu giudicata quella meno luminosa, preceduta da *dark sound of battle* (suono oscuro della battaglia).

Anche nel caso di associazioni tra suoni e stimoli visivi colorati (e quindi più complessi rispetto a immagini in scala di grigio o semplici fonti luminose) il paradigma viene confermato [143]: pitch elevati sono correlati a colori con alti valori di brillanza (*lightness*).

1.4.2 Pitch – Dimensione delle forme

In questa classe di corrispondenze cross-modali si pone a confronto l'altezza degli stimoli uditivi e la dimensione delle forme visive ad essi associate. Gli oggetti più piccoli sono generalmente abbinati ai suoni più acuti, mentre gli oggetti più grandi sono congruenti ai suoni più gravi [86][103]; tale corrispondenza sembra essere bidirezionale, ovvero ugualmente consistente partendo sia dagli stimoli acustici sia da quelli visivi [35]. L'elaborazione cross-modale tra il pitch e la magnitudine delle forme rivela così un'associazione inversamente proporzionale tra queste due *features*.

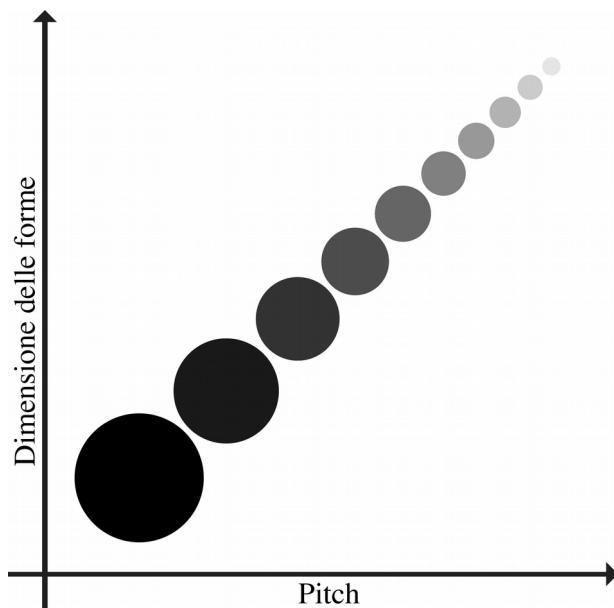


Illustration 4: Modello base per la corrispondenza cross-modale pitch-dimensione.

Tale corrispondenza sembra riflettere la naturale correlazione tra le proprietà fisiche degli oggetti del mondo esterno. La dimensione associata al pitch potrebbe infatti rispecchiare le proprietà della risonanza acustica, secondo cui i corpi più grandi risuonano a frequenze più basse rispetto ai corpi più piccoli [29][55]. Dagli *speed test* si evince che la discriminazione tra coppie di stimoli congruenti (suoni acuti – forme piccole; suoni gravi – forme grandi) avviene più velocemente rispetto alle coppie di stimoli incongruenti [46].

Di fatto però i rimandi al mondo fisico, che maturerebbero con l'esperienza in rapporto ad esso, potrebbero non essere la causa principale della

natura di tale corrispondenza. È stato infatti evidenziato come già i bambini di quattro mesi tendano ad essere più attratti da stimoli audio-visivi congruenti (pitch elevati sono in sintonia alle forme più sottili) rispetto a stimoli incongruenti [32]. La presenza precoce di tali associazioni cross-modali ha portato alcuni ricercatori a concludere che queste corrispondenze siano funzioni innate nell'attività celebrale, basate su particolari codifiche neuronali.

Va sottolineato che la quasi totalità dei test percettivi volti a valutare la correlazione tra l'altezza dei suoni e le dimensioni delle forme concorrenti hanno fatto uso di stimoli stazionari e invariabili (altezza e dimensioni fisse). Si riporta invece un caso relativamente recente in cui i soggetti sono stati sottoposti a stimoli cross-modali variabili [34]. Nel test effettuato all'università di Tel-Aviv da Eitan, Schupak, Gotler e Marks, a 64 partecipanti sono stati presentati suoni sinusoidali con aumento e diminuzione continua di altezza (glissandi ascendenti e discendenti di un'ottava) assieme a cerchi geometrici di diametro variabile in modo altrettanto continuo; ai soggetti veniva chiesto di discriminare il più rapidamente possibile (*speedtest*) la direzione del cambiamento di magnitudine delle due *features*. Il risultato ottenuto stabilisce un paradosso rispetto ai risultati dei test con stimoli stazionari: all'aumentare del pitch, infatti, i soggetti associano un aumento delle dimensioni geometriche, in netto contrasto con gli abbondanti risultati che vedono un rapporto inversamente proporzionale tra le due *features*. Gli autori del test spiegano questo insolito risultato avanzando l'ipotesi che, nel caso di tali stimoli dinamici, subentri un'associazione tra il cambiamento delle dimensioni e la variazione della distanza di osservazione. Una continua variazione della dimensione dell'immagine può effettivamente essere associata al movimento di avvicinamento o allontanamento, dal momento che gli oggetti in avvicinamento producono un ingrandimento dell'immagine retinica, mentre gli oggetti che si allontanano provocano un suo restringimento. Si tenga inoltre presente che l'osservazione di sorgenti sonore in movimento è soggetta al cosiddetto *effetto doppler*, grazie al quale viene a generarsi un'apparente variazione di altezza del suono: aumenta se sorgente e osservatore si avvicinano, diminuisce nel caso opposto. Un'ulteriore spiegazione all'insolito risultato del test percettivo viene avanzata sempre dagli autori, i quali ritengono plausibile un'analogia più astratta che vedrebbe l'aumento-diminuzione del pitch e la crescita-restringimento dell'oggetto entrambi associati a una variazione di magnitudine complessiva (alto/grande – basso/piccolo).

1.4.3 Pitch – Elevazione spaziale delle forme

Pitch ed elevazione nello spazio fu una tra le prime coppie di *features* ad essere indagata nel campo delle corrispondenze cross-modali. Il filosofo e psicologo tedesco Carl Stumpf⁷ – il quale ha a lungo indagato il suono e il suo rapporto con la psiche umana – fu il primo ad osservare come in diverse lingue vengono utilizzate le etichette “alto” e “basso” in riferimento all’altezza del suono [132]. Bernstein e Edelstein, autori del primo test nell’ambito della moderna psicologia cognitiva, verificarono che si è portati a identificare come congruenti gli stimoli visivi proiettati verso il basso ed i segnali acustici a bassa frequenza, così come le immagini spazialmente elevate se accompagnate da segnali acustici ad alta frequenza [9][10]. Viceversa, la compresenza di suoni gravi e immagini spazialmente elevate e quella di suoni acuti e immagini proiettate verso il basso non vengono altrettanto percepite come congrue. Dunque il modello che esplicita il rapporto tra pitch ed elevazione spaziale è quello mostrato nel seguente grafico:

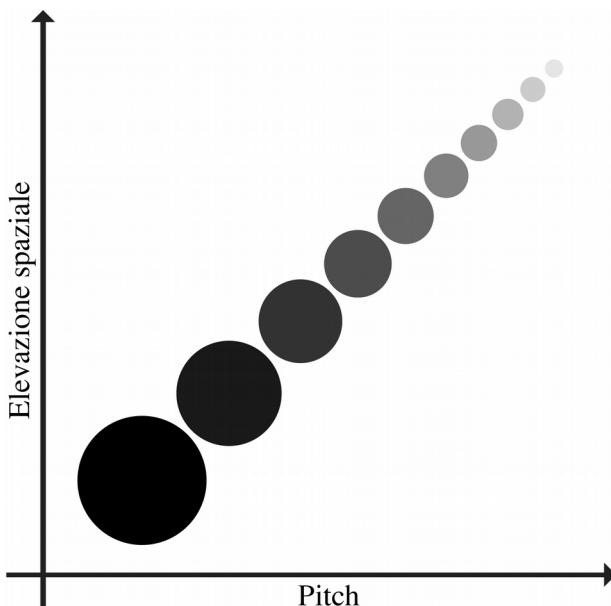


Illustration 5: Modello base per la corrispondenza cross-modale pitch-elevazione.

⁷ Si ricorda che Stumpf fu una figura molto influente nel campo della psicologia: fondatore della cosiddetta “scuola di Berlino”, i suoi studi contribuirono alla nascita della successiva psicologia della Gestalt (Köhler, Koffka e Wertheimer) e della corrente fenomenologica di Edmund Husserl, il quale fu allievo di Stumpf.

Un test di Mondloch e Maurer dimostra che bambini di trentacinque mesi sono portati ad osservare per tempi più lunghi una palla che si sposta verso l'alto se tale movimento viene accompagnato da un segnale acustico ad altezza crescente rispetto ad uno con altezza decrescente [94]. Melara e O'Brien misurarono i tempi di reazione dei soggetti partecipanti ad uno *speed test* da loro coordinato, quantificando l'effetto congruenza in tale binomio di *features* audio-visive [92]; i tempi di risposta erano pari a una media di 425 ms quando il tono e la verticalità delle forme combaciavano con il modello sopraesposto, mentre si allungavano a 450 ms quando pitch e posizione erano presentati in modo incongruo. La differenza di 25 ms in tale test rappresentava dunque la magnitudine dell'effetto congruenza. Tale dato venne ulteriormente confermato anche da un successivo *speed test* proposto da Ben-Artzi e Marks [8].

I soggetti sono inoltre più veloci nel giudicare congruenti o meno tale categoria di stimoli cross-modali quando essi vengono presentati assieme, mentre sono più lenti se gli stimoli sono presentati da soli, cioè senza il corrispettivo stimolo cross-modale [35]; questo è un ulteriore elemento che confermerebbe l'effetto di congruenza cross-modale tra il pitch e l'elevazione spaziale delle forme visive.

A livello linguistico, com'è già stato notato precedentemente, è consuetudine definire i suoni acuti come “alti” e quelli gravi come “bassi”; sorge dunque spontaneo attribuire a questa classe di corrispondenze cross-modali una implicita derivazione linguistica, una metafora spaziale del suono. Tuttavia è stato dimostrato che le mappature cross-modali tra pitch e posizione spaziale appare già in età pre-linguistica (quattro mesi)⁸ [32]; tale dato escluderebbe, o quantomeno limiterebbe, l'interferenza della lingua.

Così com'è stato evidenziato nel precedente paragrafo attinente alla corrispondenza tra pitch e dimensione delle forme, anche in questo caso la correlazione tra pitch ed elevazione spaziale potrebbe derivare da un parallelo tra la magnitudine delle due *features* [106].

Uno studio ha inoltre dimostrato su base statistica come in natura i suoni a medio-alta frequenza (soprattutto quelli compresi tra 1 e 6 kHz) siano prodotti da sorgenti sonore spazialmente elevate, e che le proprietà di filtraggio dell'orecchio umano (*Head Related Transfer Function*, HRTF) – che faciliterebbero la localizzazione delle sorgenti sonore – riflettono tale andamento ambientale; il sistema uditivo riesce facilmente a discriminare l'elevazione dei suoni ad alta frequenza, e ciò

⁸ È stato valutata la soglia di attenzione rivolta dai bambini nei confronti di stimoli audio-visivi coerenti e non; i bambini hanno mostrato più interesse per gli stimoli coerenti.

suggerisce che la corrispondenza cross-modale tra pitch ed elevazione spaziale poggi anche su una base fisiologica e ambientale [105].

Infine, test recenti definiscono questa classe di corrispondenze cross-modal come *Spatial-Musical Association of Response Codes* (effetto SMARC) [112], o anche come *Spatial-pitch association of Response Codes* (effetto SPARC) [74]; è stato dimostrato che la posizione spaziale delle forme non è variabile solamente nell'asse verticale, bensì anche in quello orizzontale: suoni acuti sono infatti congruenti con immagini posizioante in alto a destra, mentre quelli gravi sono meglio rappresentati da immagini posizionate in basso a sinistra [28].

1.4.4 Pitch – Geometria delle forme

Un’ulteriore *feature* visiva che risulta essere legata all’altezza del suono è il profilo geometrico delle forme: pitch elevati vengono correlati a forme angolari e appuntite, mentre pitch bassi richiamano forme smussate e arrotondate [92]. Gli *speed test* confermano questa corrispondenza, evidenziando tempi di reazione più veloci e precisi nella valutazione delle coppie di stimoli congruenti [81].

Dati analoghi possono essere estrapolati dagli studi sul simbolismo fonetico, volti ad approfondire il rapporto tra i suoni e il significato ad essi attribuiti. Alcune relazioni tra suono e relativo significato non sarebbero del tutto arbitrarie e in più potrebbero essere universalmente condivise; ad esempio, è stato constatato che attraverso le lingue, gli oggetti più grandi o più scuri tendono ad essere nominati da parole contenenti le vocali /a/ e /o/ [31], dato che può ulteriormente interessare la corrispondenza tra il pitch e le dimensioni delle forme. È importante rimarcare il fatto che il simbolismo fonetico (in accordo con le ricerche sulla prosodia) generalmente attribuisce un peso maggiore all’articolazione vocalica rispetto all’articolazione consonantica [100], in quanto le vocali, nella lingua parlata, sono altamente soggette alla modulazione della frequenza fondamentale (F0) del tono lessicale, rendendo variabile il significato fonetico di una medesima parola⁹ [47][123].

Il gestaltista Wolfgang Köhler propose una gerarchizzazione dei suoni vocalici sulla base del loro contenuto spettrale [69]; egli creò così una sorta di “scala vocalica” [87] con il seguente ordine di altezze crescenti: u, o, a, e, i. Tale scala vocalica è confermata dagli studi di fonetica acustica; la tabella n. 2 mostra l’altezza media delle prime due formanti delle sette vocali italiane (/e/ e /o/ vengono separate tra aperte e chiuse).

⁹ L’intonazione del parlato può suscitare diversi effetti psicologici: un tono grave solitamente induce un senso di rilassatezza e tende a mettere a proprio agio l’ascoltatore, mentre un’intonazione spinta verso il registro acuto può provocare uno stato di allerta o di fastidio.

| <i>vocale</i> | <i>F1 – F2 [Hz]</i> | <i>F1 – F2 [Hz]</i> |
|---------------|----------------------|-----------------------|
| | <i>voce maschile</i> | <i>voce femminile</i> |
| [i] | 280 – 2240 | 320 – 2750 |
| [e] | 360 – 2040 | 400 – 2500 |
| [ɛ] | 560 – 1840 | 620 – 2400 |
| [a] | 800 – 1280 | 920 – 1400 |
| [ɔ] | 520 – 900 | 640 – 1200 |
| [o] | 420 – 800 | 400 – 920 |
| [u] | 280 – 720 | 360 – 760 |

Table 2: Altezza delle vocali italiane; estratto dagli studi di Ferrero e Magno Caldognetto [40].

Gli studi di simbolismo fonetico hanno dimostrato come le vocali con formanti più gravi (/u/, /o/, e /a/) vengano generalmente associate a figure geometriche stondate (ellissi, cerchi) mentre le vocali con formanti più acute (/e/ ed /i/) sono correlate a figure spigolose (triangoli) [134].

Si noti altresì che tali studi considerano equivalenti (e quindi usati in modo intercambiabile) i suoni sinusoidali e i suoni vocalici, sebbene ciò comporti un certo grado di semplificazione [78]. Robert Tarte fu il primo ad utilizzare suoni puri sinusoidali al posto dei suoni vocalici in un suo studio¹⁰ per verificare la corrispondenza tra l'altezza dei suoni e il profilo geometrico delle forme. Ai partecipanti del suo test venne chiesto di associare una coppia di immagini (un triangolo e un'ellisse) ai suoni che ritenevano più congrui scelti tra un set di nove segnali acustici a diverse altezze e durate. Il dato più saliente consistette nel rilevare che i segnali più gravi venivano associati alla figura ellittica, mentre quelli acuti a quella triangolare.

Nel test sul simbolismo fonetico condotto da O'Boyle e Tarte per esaminare la relazione tra suoni puri e figure geometriche, il set di figure era dato e le stesse figure venivano proposte con due diverse dimensioni (piccole e grandi), mentre la corrispettiva frequenza del suono sinusoidale era a discrezione dei soggetti sottoposti al test, i quali potevano liberamente variarla tra 40 e 12250 Hz finché non risultasse percettivamente adeguata alla rispettiva figura [100]. Dal test si evince che alle forme appuntite vengono attribuite frequenze più acute, mentre quelle arrotondate sono meglio rappresentate dai suoni più gravi. L'unico dato in controtendenza è legato figura “uloomu” (vedi fig. 6), la quale, sebbene sia geometricamente costituita da una

¹⁰ Tarte, Robert D. *Phonetic symbolism for pure tones*. Paper presentato al meeting annuale organizzato dalla Psychonomic Society nel Novembre del 1976 a St. Louis, MO.

combinazione di tre figure ellittiche, ha mostrato risultati simili alla figura “takete” e in un test è addirittura risultata la forma associata alla frequenza più acuta.

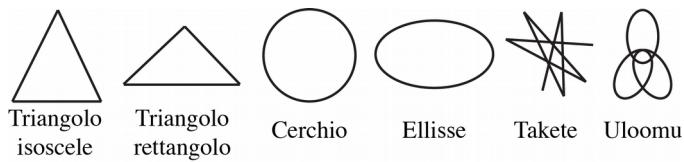


Illustration 6: Figure geometriche utilizzate nel test di O'Boyle e Tarte.

Un altro test di Parise e Spence sulle corrispondenze crossmodali audio-visive sottoponeva i partecipanti a due coppie di stimoli, una coppia di suoni sinusoidali (300 e 4500 Hz) e una coppia di figure angolari (un angolo acuto di 42° e uno ottuso di 126°) [104]; veniva poi chiesto di associare ogni immagine ad un rispettivo suono. Dalle scelte dei partecipanti del test, l'angolo acuto risulta congruente con la sinusoide ad alta frequenza, quello ottuso con la sinusoide a bassa frequenza.

Mettendo assieme i risultati dei vari test, un possibile modello esplicativo per la corrispondenza tra pitch e geometria è il seguente:

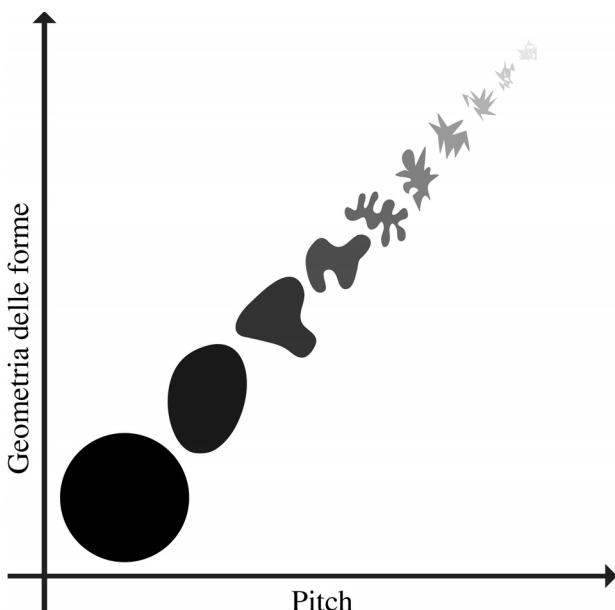


Illustration 7: Modello base per la corrispondenza cross-modale pitch-geometria.

1.4.5 Loudness – Luminosità

La loudness e la luminosità condividono la stessa variabile fisica, l'intensità, ossia il grado di forza con cui si manifestano i fenomeni della luce e del suono. Se si volesse avanzare un parallelo tra il suono e il colore secondo questo criterio, i colori più chiari sarebbero correlati ai suoni forti (il bianco come caso limite), mentre i colori scuri andrebbero d'accordo con i suoni più deboli (il nero equivale al silenzio) [23]. Tale corrispondenza risulta consistente nei soggetti sinestetici [78], per cui, stando alla somiglianza con i processi cross-modali, tale binomio dovrebbe essere riscontrato anche nei processi cognitivi condivisi da tutti.

Marks e Stevens nel 1966 dimostrarono l'effettiva validità di tale funzione cross-modale tra le due *features* [76] in un test nel quale i partecipanti dovevano settare autonomamente l'intensità di una banda di rumore (75-4800 Hz) e il livello di luminanza della luce di un proiettore; i livelli di loudness e di luminanza venivano regolati in maniera proporzionale dagli stessi partecipanti al test, così come viene mostrato in figura 8.

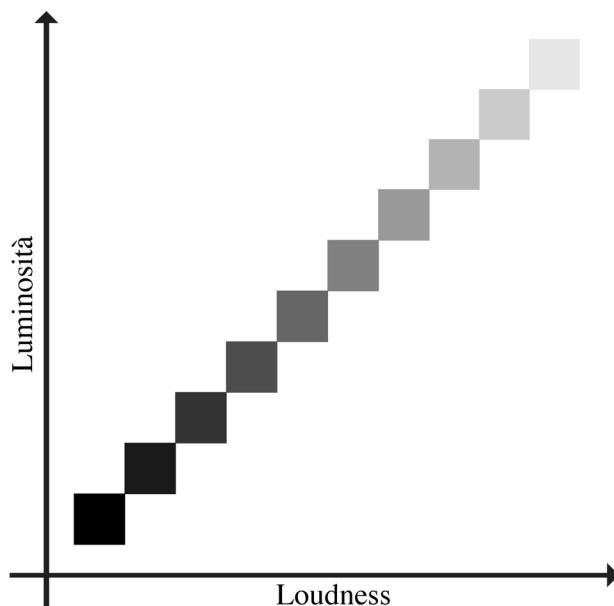


Illustration 8: Modello base per la corrispondenza cross-modale loudness-luminosità.

Uno studio di un anno più tardi ottenne dei risultati non altrettanto soddisfacenti. Tulving e Lindsay nel 1967 coordinarono un test [137] nel quale i soggetti venivano

sottoposti a stimoli visivi (cerchi luminosi con vari livelli di luminanza) e stimoli uditivi (sinusoidi a 1000 Hz a varia intensità) di varia durata (20 ms, 50 ms e 2 s). Il test si sviluppava in tre prove differenti: nelle prime due veniva chiesto ai partecipanti di classificare su una scala da 1 a 8 gli stimoli presentati singolarmente e con le durate più brevi (20 e 50 ms), senza dunque uno stimolo concorrente nel secondo canale percettivo; nella terza prova gli stimoli venivano proposti simultaneamente e con durata più lunga (2 s), e anche in questo caso veniva chiesto di classificare la coppia di stimoli in base alla loro reciproca congruenza. La premessa dell'esperimento poggiava nell'ipotesi che stimoli molto brevi non darebbero ai soggetti il tempo necessario per elaborare e classificare in modo sicuro l'entità della *feature* da valutare, mentre tempi più lunghi e, in aggiunta, accompagnati da un secondo stimolo nel quale venga variato in modo simmetrico un suo tratto presumibilmente corrispondente, faciliterebbe tale classificazione. Se da una parte il test dimostra che tempi più lunghi degli input visivi e uditivi migliorino il trasferimento dei loro dati percettivi, dall'altra l'unione di stimoli appartenenti a due modalità diverse non evidenzia prestazioni migliori; la prestazione complessiva infatti, misurata in termini di informazioni trasmesse, era essenzialmente la stessa in presenza e in assenza di stimolazione cross-modale. Un simile risultato farebbe pensare che la corrispondenza cross-modale tra loudness e luminosità non sia molto robusta come quelle viste nei precedenti paragrafi. Il fatto che in tale test fossero coinvolte solamente quattro persone non può comunque muovere un giudizio assoluto su tale classe di corrispondenze; in aggiunta, successivi test ottennero dei risultati contrastanti rispetto a quelli di Tulving e Lindsay. Innanzitutto è stato mostrato come i bambini di cinque anni tendano ad abbinare la luminosità di una luce all'intensità di un suono [14], e successivamente come negli *speed test* la simultaneità di stimoli visivi e uditivi incongruenti generi un effetto di interferenza che rallenta i tempi di reazione [81].

Ad ogni modo, in questa classe di corrispondenze cross-modali la relazione percettiva tra le due *features* non è così marcata come, ad esempio, quella tra il pitch e la luminosità [81][82]. Lawrence Marks propose una nuova *feature* percettiva del suono a cui diede il nome di *densità*, la quale sarebbe definibile da una sommatoria di loudness e pitch [78][82][116][130]; infatti le due corrispondenze cross-modali facenti capo alla luminosità potrebbero di fatto essere accorpate in una superclasse che vedrebbe da una parte la luminanza visiva e dall'altra la densità uditiva, dove, da un'analisi multidimensionale delle risposte sulla corrispondenza tra densità e luminosità, si è tuttavia compreso che la magnitudine del pitch gioca un peso maggiore rispetto a quella della loudness [82].

1.4.6 Loudness – Dimensione delle forme

La corrispondenza cross-modale tra la loudness e la dimensione delle forme è stata riscontrata da Robert Walker in un ampio ventaglio di partecipanti di età diversa e provenienti dai contesti culturali e linguistici più vari [142]. È ampiamente condivisa l'associazione di suoni deboli a forme di ridotte dimensioni, e di suoni forti a immagini più grandi [75][133]. Già a due anni i bambini mostrano di considerare congruenti stimoli uditivi ad alta intensità e forme di grandi dimensioni [125].

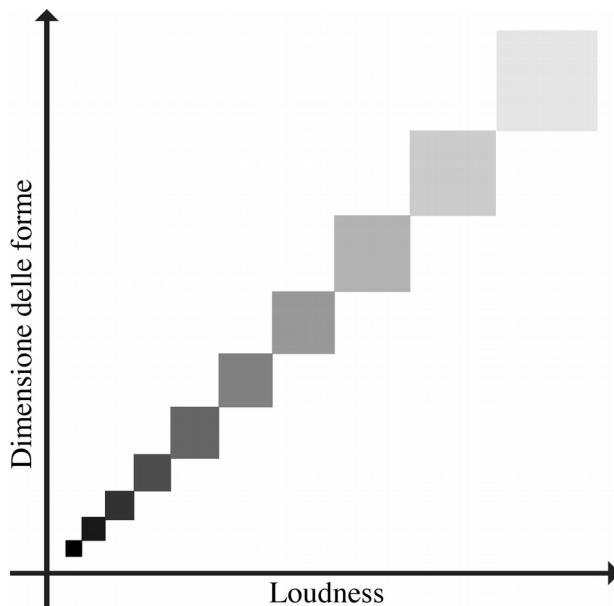


Illustration 9: Modello base per la corrispondenza cross-modale loudness-dimensione.

L'eminente psicofisico Stanley Stevens analizza la relazione percettiva tra loudness e pitch, e considera il volume una *dimensione fenomenica* del suono in grado di farlo apparire più o meno grande [131]; così, ogni tono può avere diverse *dimensioni volumiche* entro un continuum. È dello stesso parere Lawrence Marks, secondo il quale i suoni differiscono non solo nella loro intensità soggettiva (loudness), ma anche nel grado in cui sembrano riempire lo spazio. I suoni forti e gravi appaiono più massicci e sembrano occupare un grande volume, mentre i suoni tenui e acuti appaiono più sottili, piccoli e sembrano occupare un piccolo volume [79].

Anche la corrispondenza cross-modale tra loudness e dimensione potrebbe risiedere su correlazioni esperienziali tra la dimensione degli oggetti e il volume dei suoni che tendono a produrre. Infatti, nell'esperienza uditiva di tutti i giorni (*everyday listening*) la loudness ha un ruolo prominente nella discriminazione uditiva della lunghezza e della dimensione degli oggetti [25].

4.7 Timbro – Geometria delle forme

Quest'ultima classe di corrispondenze cross-modali ad oggi conta un esiguo numero di studi, ma qualche elemento per approfondire la loro corrispondenza può essere tirato in causa.

In primo luogo è possibile trarre dei parametri di valutazione nuovamente dagli studi sul fono-simbolismo. In tale campo di indagini, una lunga serie di test percettivi deriva dalla nota contrapposizione proposta da Wolfgang Köhler tra una coppia di figure (una dai contorni tondeggianti e l'altra dai contorni spigolosi) e una coppia di parole (*maluma* e *takete*) [70]; nel test di Köhler ai soggetti veniva chiesto di associare ogni parola a una figura corrispondente. Fu sorprendente rilevare la naturalezza con la quale in modo unanime venisse associato *maluma* alla figura rotondeggiante e *takete* a quella spigolosa.

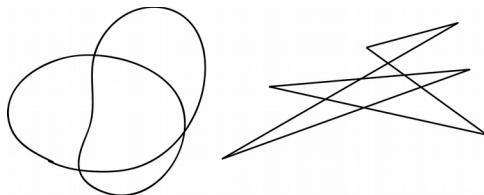


Illustration 10: *Maluma e Takete*.

Ramachandran e Hubbard proposero un test simile nel 2001 utilizzando le pseudo-parole *bouba* e *kiki*, e rilevarono che la maggior parte dei partecipanti (circa il 98%) associava *bouba* a una forma sinuosa, ed una forma angolare a *kiki* [111]; questa corrispondenza cross-modale è diventata nota con il nome di *bouba-kiki effect*, e la sua validità è stata riscontrata in vari contesti linguistici [111][135], interculturali [30] [20] e per bambini in età pre-linguistica [63][90][101].



Illustration 11: *Bouba e Kiki*.

Se, come si è visto precedentemente, i suoni vocalici giocano un ruolo di primaria importanza per la modulazione dell'altezza fonetica, anche le consonanti possono

contribuire alla correlazione di talune pseudo-parole a tal altre forme geometriche. Per esempio, nelle parole takete e maluma sono chiare le differenze acustiche dovute ai transitori d'attacco e alla distribuzione spettrale dei gruppi di suoni consonantici /t/ /k/ e /m/ /l/: le prime vengono percepite come consonanti “dure”, mentre le seconde appaiono più “morbide”. Uno spettro sonoro con transienti più ripidi e ricco di alte frequenze è così associato a forme appuntite, mentre uno spettro con baricentro sul registro medio-basso e con un attacco dolce è rappresentato con curve e profili smussati. Appositi studi di approfondimento sulle corrispondenze visive attribuite ai diversi ceppi di suoni consonantici hanno evidenziato che le consonanti stoppate (b, k, p, t) vengono rapidamente associate a geometrie angolari, mentre quelle continue (f, m, n, r) alle forme arrotondate [97][98][147]. Tale associazione è stata attribuita da alcuni ad una co-attivazione – di tipo pseudo-sinestetico – dell’area cerebrale somatomotoria e della corteccia visiva [111]; in questo senso, la parola *maluma*, che richiede un certo tipo di articolazione labiale (l’arrotondamento della bocca), richiama involontariamente delle figure ellittiche e circolari.

Uscendo dal campo del simbolismo fonetico, Parise e Spence, attraverso un apposito *speed test*, verificarono che un’onda quadra viene percepita congruente ad una geometria angolare, mentre un suono sinusoidale è rappresentato da una figura curvilinea [104].

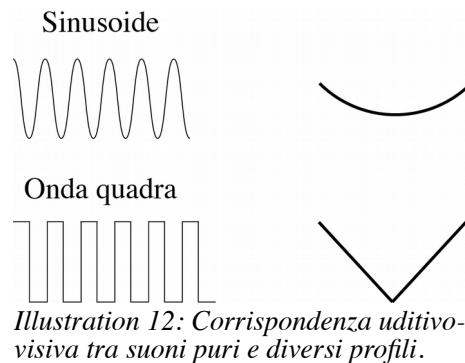


Illustration 12: Corrispondenza uditivo-visiva tra suoni puri e diversi profili.

Un test percettivo coordinato da Lipscomb e Kim chiedeva di associare tre suoni aventi stessa altezza e stessa loudness ma prodotti da strumenti diversi (flauto, corno francese e tromba) a tre figure geometriche differenti (un cerchio, un quadrato e un triangolo); i partecipanti, tra i quali vi erano sia soggetti con esperienza in ambito musicale sia soggetti senza alcuna preparazione in merito, assocavano il suono di flauto alla figura circolare, il suono di corno inglese al quadrato e il suono di tromba al triangolo [75]. Un ulteriore test che ha approfondito la corrispondenza tra timbro e

profilo geometrico delle forme è quello recentemente proposto da Adeli, Rouat e Molotchnikoff [1]; anche in questo caso sono stati impiegati timbri complessi, poiché gli stimoli uditi consistevano in campionamenti di otto strumenti musicali diversi (piano elettrico, marimba, chitarra classica, violoncello, sax tenore, triangolo, crash e gong) trasposti in quattro differenti altezze (chiaramente solo quelli intonabili, a 100, 150, 200 e 250 Hz). Dai risultati del test si evince che vi è una forte associazione tra i timbri “morbidi-levigati” (soprattutto piano elettrico e marimba) e forme arrotondate, mentre ai timbri “duri-aspri” vengono associate forme appuntite; nella fattispecie, i timbri corrispondenti alle forme più appuntite risultano essere quelli degli strumenti percussivi (triangolo, crash e gong). Anche qui sembra determinante il profilo d’attacco del suono.

Dunque il timbro sembrerebbe anch’esso modulare i tratti delle corrispondenze visive. La percezione qualitativa del timbro è tuttavia descrivibile da più *features* percettive di basso livello, come ad esempio lo *spectral centroid* [138] che a sua volta determina l’*indice di chiarezza (brightness)*, ma anche dalla *sharpness* [13], la *roughness* [36] e da altre ancora. Sarebbe dunque utile continuare ad indagare per comprendere quali dimensioni percettive timbriche di basso livello siano determinanti nella modulazione delle corrispondenze visive.

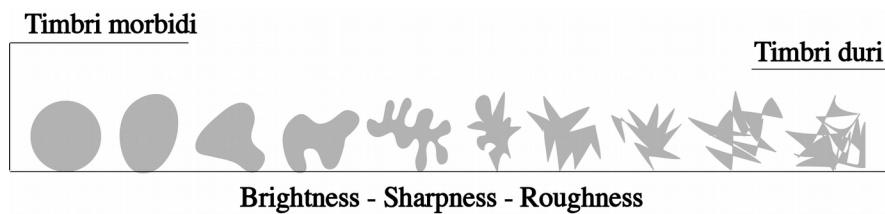


Illustration 13: Modello base per la corrispondenza cross-modale fra timbro sonoro e profilo della forma.

1.5 Illusioni percettive

È stato dimostrato come la percezione cross-modale in certe condizioni possa essere alterata da particolari effetti illusionistici [119]. Ad esempio, il cosiddetto *effetto McGurk* è il risultato di una integrazione cross-modale falsata: i soggetti sottoposti allo stimolo acustico /ba/ e simultaneamente ad uno stimolo visivo che mostra una persona mimare il fonema /ga/ sono portati ad integrare – e dunque percepire – un terzo fonema /da/, frutto di una storpiatura dei dati cross-modali [91]. Similmente, persone a cui viene fatto udire il suono “baba” sincronizzato ad un video di un attore che articola la parola “gaga”, tendono a combinare le due informazioni (uditiva e visiva) nel fonema “dada”.

Nei processi di convergenza cross-modale la localizzazione spaziale dei segnali acustici attiva l’interazione del sistema uditivo e di quello visivo. In alcuni casi gli stimoli in arrivo alle diverse modalità sensoriali forniscono informazioni contrastanti sulla medesima caratteristica di un singolo evento. Per quanto concerne la localizzazione della sorgente acustica, l’*effetto ventriloquo* [12] implica un conflitto percettivo tra il dato uditivo e quello visivo, guidato dalla localizzazione visuo-saziale¹¹.

Tuttavia vari studi, indagando sulle corrispondenze di pattern temporali tra la modalità uditiva e quella visiva, hanno dimostrato la capacità del suono nell’alterare la corretta percezione del dato visivo. È chiaro, nel caso di flussi audio-visivi in rapido movimento, la qualità della percezione cross-modale migliora se vi è sincronismo tra i due canali sensoriali [43], sincronismo che facilita l’individuazione dei singoli eventi in modo puntuale grazie alla ridondanza dei dati percettivi [140]. Prese singolarmente, la percezione uditiva [17] e quella visiva sono in grado di segmentare in singoli oggetti discreti i rispettivi pattern temporali di dati in arrivo (*stream segmentation*) entro una certa soglia di frequenza.

Si prenda l’esperimento di Bregman e Achim [18] dove veniva mostrato a monitor un array di quattro punti luce, illuminati in modo alternato (in sequenza 1, 2, 3, 4, vedi figura 14) a varie velocità (1, 2 e 4 sequenze al secondo). A basse velocità (1-2 Hz) la segmentazione visiva (*visual stream segmentation*) veniva percepita in modo corretto, cioè i soggetti erano in grado di distinguere chiaramente le singole

¹¹ Ciò conferma peraltro la superiorità dell’apparato visivo sugli altri sensi nei processi percettivi di base.

sorgenti luminose, mentre a velocità più alta (4 Hz) i soggetti tendevano a percepire due flussi visivi sovrapposti, così com'è mostrato in figura 14.

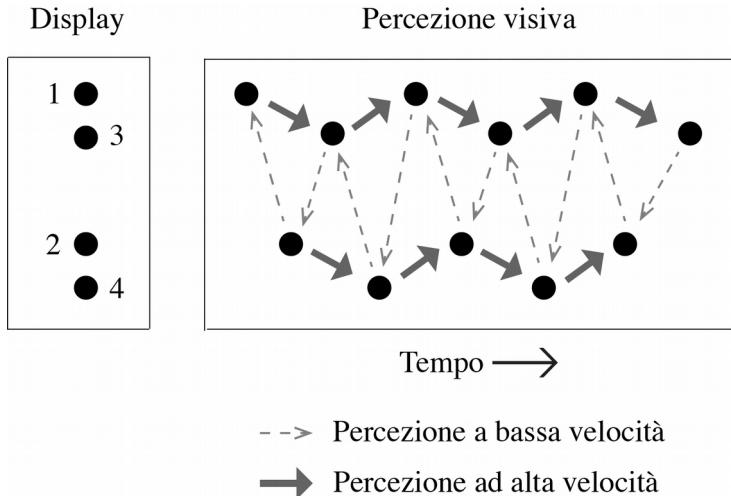


Illustration 14: Segmentazione visiva con quattro punti luminosi, illuminati alternativamente in quest'ordine: 1, 2, 3, 4 a bassa e alta velocità.

In modo analogo, un pattern uditivo composto da una continua alternanza di brevi impulsi sinusoidali (100 ms) rispettivamente ad alta e bassa frequenza viene segmentato correttamente (*auditory stream segmentation*) se l'alternanza tra i toni acuti e gravi avviene a bassa frequenza (3 Hz), ma, al contrario, vengono percepiti due flussi continui e sovrapposti qualora l'alternanza avvenga ad una frequenza poco maggiore (5-10 Hz) [19].

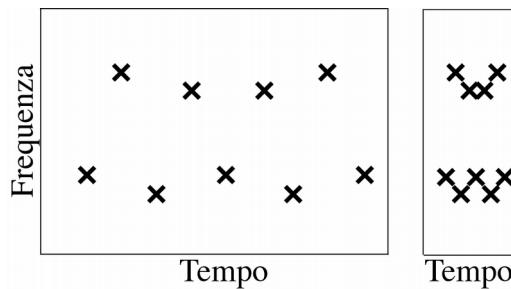


Illustration 15: Differenti segmentazioni dello stream audio a bassa e alta frequenza.

La segmentazione percettiva in una modalità può essere influenzata dalla simultanea stimolazione covariante in un'altra modalità [99]. Nel caso di eventi puntuali, per singole coppie di stimoli uditivi e visivi, la presenza di suono può indurre a percepire in modo più accentuato la luminosità di uno stimolo visivo, in rapporto allo stesso input visivo privo di suono [128]. È stato inoltre evidenziato come la percezione di un

singolo flash di luce possa venire falsata qualora esso sia accompagnato da molteplici impulsi sonori; in questo caso l'individuazione del pattern uditivo condiziona il dato visivo, illudendo la simultanea percezione di un equivalente treno di impulsi visivi (*sound-induced illusory flashing*) [118][119].

Un test eseguito nel 1964 da Thorne Shipley, sottopose dei soggetti a treni di impulsi sonori (click) e treni di impulsi visivi (flash di luce); partendo da una situazione iniziale di perfetto sincronismo tra i due stimoli (entrambi a 10 Hz e in fase tra loro), venne evidenziato come un aumento o una diminuzione di frequenza degli impulsi uditivi induce la percezione di una equivalente variazione sincrona degli stimoli visivi, sincronismo che tra i due stimoli non sussiste affatto [121]. Ciò è in accordo con gli studi di Welch, DuttonHurt e Warren, i quali hanno rilevato che, nella ricostruzione percettiva di pattern temporali audio-visivi a bassa frequenza (tra 4 e 10 Hz), l'audizione esercita un peso maggiore rispetto alla visione [145]. Al contrario, una variazione nella frequenza degli impulsi visivi non induce una altrettanto equivalente variazione percettiva del pattern sonoro. Si può dunque sostenere che la percezione uditiva, a determinate condizioni, possa guidare e alterare quella visiva.

In questo breve scorcio sono stati riportati alcuni casi particolari che dimostrano come le funzioni cognitive cross-modali possano essere tratte in inganno. Il sistema percettivo scandaglia di continuo l'ambiente circostante e nel caso in cui esso rilevi una serie di stimoli uditivi e visivi prossimi tra loro, l'integrazione cross-modale tenderà a tradurre i due dati sensoriali come se fossero originati da un'unica fonte (*crossmodal binding*); nel caso particolare in cui gli stimoli uditivi e visivi si presentino sotto forma di flussi aventi delle velocità tali da rendere impossibile la loro segmentazione in singoli eventi discreti, la mente elabora delle approssimazioni nelle quali una modalità può guidare l'altra, a seconda dei casi specifici, accorpando i diversi flussi sensoriali e falsando il reale dato fisico.

1.6 Corrispondenze empiriche suono – colore

Per alcune persone le sensazioni uditive provocano un'esperienza visiva involontaria, automatica, molto precisa e in certi casi idiosincratica. Ad esempio, il Do centrale di un pianoforte può suscitare un fotismo rosso, oppure verde nel caso della stessa nota tre ottave più in alto [54]. Come è stato visto all'inizio di questo scritto, tali effetti percettivi sono tipici negli individui sinestetici, per i quali le esperienze di audizione colorata (*chromesthesia*) sono le più frequenti.

Sempre nel primo paragrafo è stato fatto un breve accenno ad alcune tra le moltissime proposte di corrispondenza suono-colore avanzate nel corso dei secoli da figure di spicco sia del mondo scientifico sia di quello artistico; è tuttavia necessario ribadire che tali modelli sono frutto di riflessioni speculative, guidate dalla soggettività dei loro autori, e nessuna di esse si è servita di un vero e proprio metodo scientifico quale strumento di supporto atto a consolidare le rispettive equivalenze suono-colore. Un'ulteriore modello di correlazione tra suono e colore ipotizzato in tempi non lontani su una base vagamente più scientifica è quello proposto da Wendell Garner [49], il quale suggerì una correlazione tra le dodici note della scala cromatica e lo spettro di colori puri percepito dall'occhio, così come mostrato in figura 16. Vista e udito infatti, secondo Garner mostrerebbero un'organizzazione per bande d'ottava simile. In un articolo dove espone la sua teoria [49], la frequenza del Do₃ di un pianoforte viene approssimata a 125 Hz e in dodici step di 10 Hz ciascuno si raggiunge per semitonni il Do₄ all'ottava superiore, con una frequenza pari a 250 Hz. Per quanto riguarda il colore, lo spettro di luce visibile è circoscritto in un range tra la lunghezza d'onda massima di 0.8×10^{-3} mm del rosso e quella minima di 0.4×10^{-3} mm del giallo; il reciproco di tali lunghezze d'onda determinerebbe la frequenza del colore nell'unità di tempo (un equivalente di fortuna alla frequenza del suono non contemplato dalla fisica del colore), ottenendo così gli stessi valori di 125 Hz per il rosso (R) e 250 Hz per il giallo (Y). In realtà il modello di Garner non si discosta molto dai precedenti sistemi proposti da Isaac Newton (c. 1700), dal matematico gesuita francese Louis B. Castel (c. 1720-1735), dall'architetto Edmund George Lind (*The music of color*, 1894) e dal compositore Edward Maryon (*Marcotone*, 1924) [146]; sebbene egli abbia cercato di giustificare con una base matematica le corrispondenze tra note e colore, vi sono molti coni d'ombra che rendono fumosa tale correlazione. Uno su tutti: l'intero spettro visibile verrebbe così coperto da una sola ottava, quando l'estensione udibile per l'uomo è di circa dieci ottave.

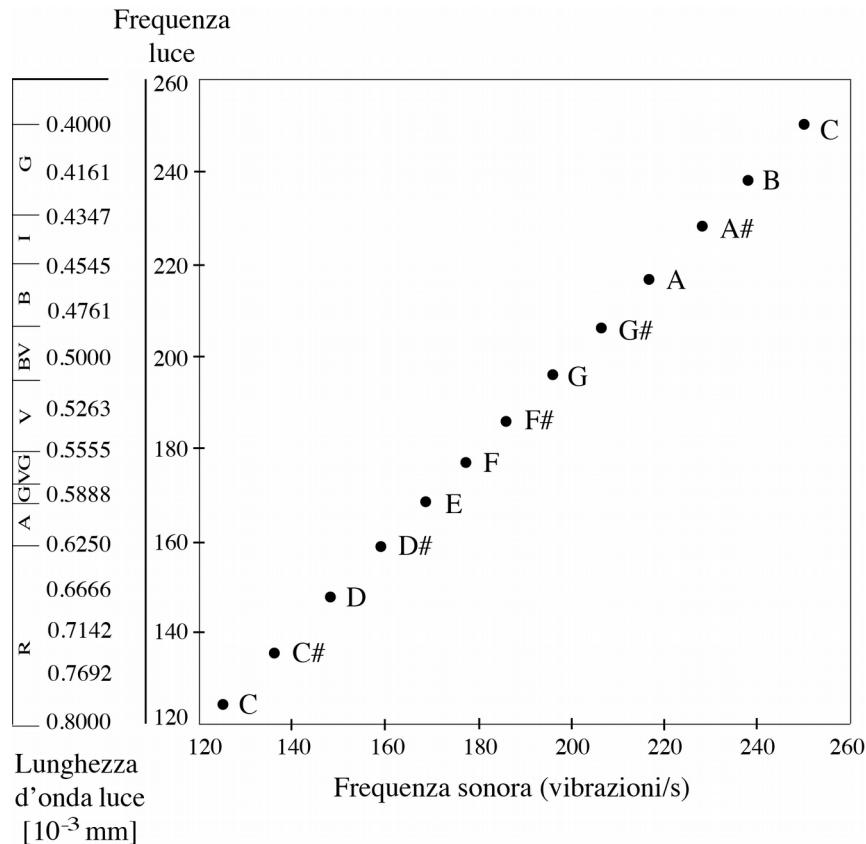


Illustration 16: Rapporto tra la frequenza sonora e la "frequenza del colore" proposto da Garner (1978).

In anni più recenti sono stati proposti dei modelli che relazionano suono e colore anche sulla base dei risultati emersi dalle ricerche cognitive sulla percezione cross-modale.

Un primo modello che viene qui riportato è quello proposto da José Luis Caivano nel 1994 [23], il quale, riprendendo da una parte i modelli suono-colore del passato e dall'altra appunto gli studi sulla percezione cross-modale, suggerisce una corrispondenza tra le feature del colore e quelle del suono, riassunte nella tabella numero 3:

| <i>Features colore</i> | <i>Features suono</i> |
|------------------------|-----------------------|
| Tonalità | Pitch |
| Luminosità | Loudness |
| Saturazione | Timbro |

Table 3: Modello suono-colore Caivano (1994)

Caivano ripropone a titolo esemplificativo tutta una serie di relazioni tra tonalità del colore e altezza delle note, proposte in passato da diversi autori, tra cui proprio il modello di Garner. Nonostante Caivano si trovi sostanzialmente in accordo con Hermann von Helmholtz – il quale a sua volta sosteneva che ogni tentativo di comparazione tra la musica e il colore dovrebbe essere abbandonato¹² – la sua posizione è un pò meno radicale; nonostante la corrispondenza tra l’altezza delle note e la tonalità del colore venga spesso ragionata in modo pretestuoso, egli sostiene che una sorta di relazione vi sia, ma essa deve essere intesa come una semplice analogia generale. In accordo invece con i risultati dei test percettivi nel campo delle corrispondenze cross-modali, Caivano suggerisce una relazione tra la luminosità del colore su scala Munsell e la loudness sonora, quest’ultima rapportata alle indicazioni di dinamica segnalate in notazione musicale tradizionale (da *ppp* a *fff*):

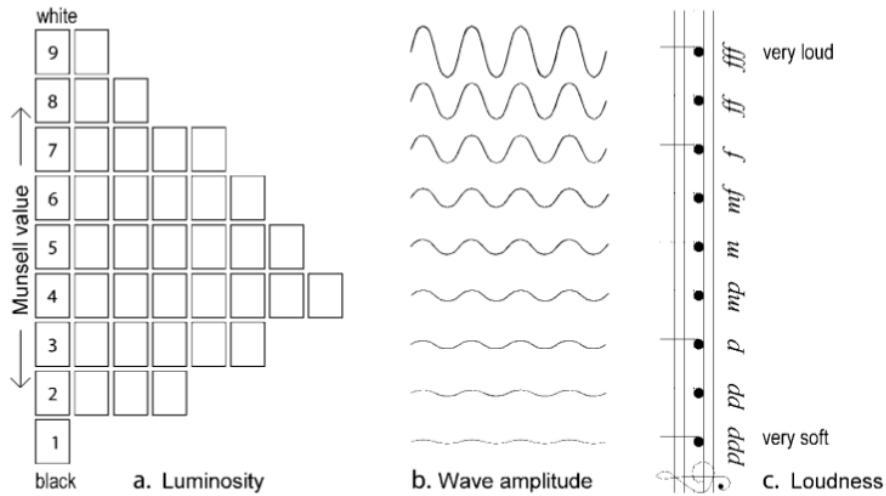


Illustration 17: Rapporto tra luminosità del colore e loudness sonora; José Caivano.

Se le relazioni tonalità-altezza e luminosità-loudness sono binomi già visti, Caivano propone una relazione nuova che mette a confronto la saturazione del colore con il timbro del suono.

Timbre is the dimension of sound that depends on the way in which the harmonics are grouped around a fundamental sound; it has to do with the quantity and relative intensity of these harmonics; we can say that this is what makes the “complexity” of the sound. Timbre is the variable by

12 *Handbuch der Physiologischen Optik (Manuale di fisiologia ottica, 1856-1867).*

which we can recognize the sound of the different musical instruments or human voices, even when they emit exactly the same tone. Timbre has been called the “color” of sound. In the musical jargon it is usual to talk of pure, rough, piercing, soft, smooth, velvety timbres, and so on¹³. [23, p.131]

Il parallelo viene retto da una sorta di “legge della purezza”: egli paragona i suoni puri con i colori puri prodotti da una banda molto ristretta dello spettro di luce che, proprio per questo motivo, possiedono la massima saturazione. Più i timbri sono complessi più, in parallelo, la saturazione diminuisce, fino al caso limite di un suono irregolare (ad es. una percussione non intonata) che corrisponderebbe così ad un colore totalmente desaturato, cioè una tonalità di grigio.

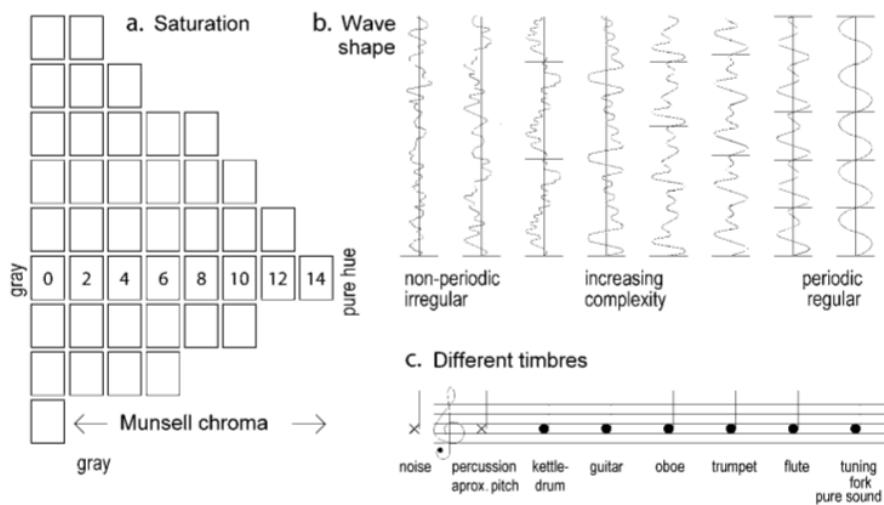


Illustration 18: Rapporto tra tonalità (chroma) del colore e timbro sonoro; José Caivano.

Tale corrispondenza può essere confutata dai risultati di un test eseguito una dozzina di anni più tardi da Ward, Huckstep e Tsakanikos al dipartimento di psicologia della University College di Londra [143], test che mostra la similitudine dei processi cognitivi tra soggetti sinestetici e non¹⁴. Come si evince dai risultati del test, viene

-
- 13 Il timbro è la dimensione del suono che dipende dal modo in cui le armoniche sono raggruppate attorno a un suono fondamentale; ha a che fare con la quantità e la relativa intensità di queste armoniche; possiamo dire che questo è ciò che rende la “complessità” del suono. Il timbro è la variabile con cui possiamo riconoscere il suono dei diversi strumenti musicali o delle voci umane, anche quando emettono esattamente lo stesso tono.
Il timbro è stato definito il “colore” del suono.
- 14 Com’è già stato sottolineato all’inizio di questo scritto, la sostanziale differenza è che i soggetti sinestetici, nel caso di audizione colorata, mostrano una reazione automatica

confermato che il timbro sonoro eserciti un effetto di congruenza in rapporto alla saturazione (*chroma*) del colore, ma, contrariamente all’ipotesi di Caivano, timbri complessi tendevano in questo specifico test ad essere associati a colori più saturi, mentre i suoni puri risultano congruenti ai colori meno saturi.

Giannakis e Smith propongono un approccio visivo alla sintesi del suono [52] retto dalla mappatura tra un sistema simbolico fondato sulle proprietà del colore e le dimensioni percettive dei suoni musicali. Le associazioni uditivo-visive derivate dagli studi sui processi cognitivi cross-modali, a detta degli autori potrebbero contribuire alla progettazione di strumenti di sintesi del suono basati sull’esteriorizzazione delle “immagini uditive interne”. Il dato più saliente che emerge da un apposito test coordinato dagli stessi Giannakis e Smith per studiare i legami percettivi tra le dimensioni sonore e visive è legato alla saturazione del colore, dimensione percettiva che risulta questa volta strettamente correlata al livello di loudness; variazioni del livello di loudness non sembrano in alcun modo interagire con la luminanza e la tonalità.

La tonalità cromatica, presa singolarmente, non risulta essere mappabile con alcuna feature uditiva, tantomeno con l’altezza del suono così come invece si è più volte cercato di dimostrare. Le varie teorie pregresse sul rapporto tra suono e tonalità del colore risultano così delle strade prive di un reale fondamento scientifico, o che quantomeno non sono sviluppate su una base psico-percettiva; esse si limiterebbero infatti a delle semplici analogie, più o meno forzate, tra la fisica del colore e quella del suono.

In chiusura di questo primo capitolo si riporta un test del tutto singolare nel campo degli studi sulla percezione cross-modale. L’esperto di neuroscienza cognitiva Jamie Ward e un team di collaboratori eseguirono una sperimentazione con l’ausilio di stimoli audio-visivi colorati realizzati seguendo le precise indicazioni di soggetti sinestetici [144]; a detta degli autori, sebbene i processi sinestetici e cross-modali in generale siano stati abbondantemente studiati, poca ricerca è stata fatta sull’influenza che essi imprimono nei giudizi estetici di gradimento o di preferenza. In tre diverse prove, un cospicuo gruppo di soggetti di controllo (ben 157) è stato sottoposto a diverse animazioni audio-visive derivate direttamente dal contributo di alcuni soggetti che presentavano forme di audizione-colorata; essi infatti guidarono, secondo la loro esperienza sinestetica, la preparazione di clip audio-video¹⁵, le quali furono successivamente presentate ai partecipanti dello studio sia nella loro forma originale,

(vedi [93]) e consistentemente uguale nel tempo agli stimoli (vedi anche [4]).

15 Le clip sono liberamente consultabili dalla piattaforma YouTube al seguente url: https://www.youtube.com/watch?v=O8Die3XX_NY; in alternativa si cerchi le parole chiave ‘*animations samantha moore*’.

sia manipolate (ruotate, con alterazione dei colori o con accoppiamento di suoni e immagini randomiche). Ai soggetti veniva dunque chiesto di giudicare secondo un indice di gradimento le varie clip (sia quelle originali che quelle di controllo), esprimendo una valutazione soggettiva sulla qualità delle corrispondenze tra la componente sonora e quella visiva. Le “animazioni sinestetiche” erano generalmente preferite rispetto a quelle di controllo (ovvero quelle manipolate). Tali risultati offrono due spunti di riflessione: in primo luogo le esperienze delle persone che presentano fenomeni di audizione colorata, se esplicitate attraverso prodotti audio-visivi, possono essere apprezzate anche dai non-sinesteti, e tale dato consoliderebbe l’ipotesi secondo cui sinesteti e non, di base condividono gli stessi processi di percezione cross-modale. Inoltre, uno studio approfondito delle congruenze sinestetiche potrebbe fornire dei criteri utili per lo sviluppo di prodotti multimediali come quelli audio-visivi.

Detto ciò, rimane tuttavia molto complicata l’individuazione di un preciso modello in grado di spiegare il rapporto tra suono e colore, quantomeno se entrambi i termini dell’equazione non vengono scissi nelle loro singole dimensioni percettive.

Capitolo 2

Percezione cross-modale: bibliografia annotata, 2016-2018

2.1 2016

S. Lacey, M. Martinez, K. McCormick, K. Sathian. *Synesthesia strengthens sound-symbolic cross-modal correspondences* [72]

Come si è visto nella prima parte di questo scritto, il confronto tra il comportamento di soggetti sinesteti e non-sinesteti è stato oggetto di numerose ricerche da parte della psicologia cognitiva e delle neuroscienze, ricerche dalle quali sono scaturite posizioni contrastanti nella stessa comunità scientifica. Per alcuni, infatti, sinestesia e percezione cross-modale rappresenterebbero i due estremi di un continuum, dove la sinestesia consisterebbe in una sorta di esasperazione degli effetti percettivi cross-modali; per altri invece, le due esperienze sono viste come processi cognitivi completamente diversi.

Al fine di porre maggiore chiarezza in questo dibattito, Simon Lacey e colleghi, attraverso un apposito *speed test*, hanno confrontato la sensibilità di sinesteti e non-sinesteti circa alcune coppie di corrispondenze audio-visive ben note, tra le quali il pitch e l'elevazione spaziale, il pitch e le dimensioni visive e infine il fonosimbolismo tra pseudo-parole e profili geometrici. L'idea alla base della ricerca era che se la sinestesia e le corrispondenze cross-modali formano un continuum, allora i soggetti sinesteti e il gruppo di controllo avrebbero mostrato risposte analoghe agli stimoli, con la sostanziale differenza che i sinesteti avrebbero collezionato le risposte più veloci data la loro maggiore sensibilità agli stimoli audio-visivi.

I soggetti sottoposti al test furono suddivisi in due gruppi: 17 individui con accertata sinestesia (previo specifico test) e 18 individui di controllo non-sinesteti.

Per verificare le classi di corrispondenze cross-modali precedentemente elencate, sono stati preparati diversi stimoli uditivi e visivi. Per quanto concerne le prove di fonosimbolismo, sono state registrate due pseudo-parole, *lomo* e *kike*; per la prima è stata prodotta un'immagine congruente dai contorni arrotondati, mentre per la seconda un'altra con un profilo appuntito. Le prove relative al rapporto tra pitch ed elevazione spaziale si sono basate su due stimoli uditivi, entrambi suoni sinusoidali della durata di 1 s, rispettivamente a 180 e 1440 Hz; gli stimoli visivi concorrenti erano due quadrati bianchi disposti diagonalmente con un vertice in comune e collocati nella parte alta o bassa del monitor con sfondo nero. Nelle prove atte a verificare le corrispondenze tra pitch uditivo e dimensioni delle forme sono stati usati

gli stessi segnali acustici, questa volta combinati con due cerchi bianchi di diverso diametro proiettati al centro del monitor (gli autori riportano le dimensioni delle figure circolari calcolate sulla base della loro copertura dell'angolo visuale¹⁶: 1° quello piccolo e 2.5° quello grande). Prima di eseguire il test, ogni partecipante impostava la loudness dello stimolo uditivo più acuto ad un livello che non gli risultasse fastidioso; l'intensità del segnale a bassa frequenza veniva poi impostato ad un livello tale da mantenere costante la loudness per entrambi i segnali acustici.

Ad ogni prova, veniva presentato uno stimolo, uditivo o visivo, e una coppia di stimoli concorrenti appartenenti alla modalità percettiva corrispondente. Ai soggetti veniva chiesto di stabilire il più rapidamente possibile a quale dei sue fosse congruente e a quale invece apparisse incongruente. L'intero test si componeva di 192 prove complessive, equamente suddivise per ambo i canali percettivi: in 96 prove gli stimoli rilevanti erano uditivi, mentre nelle altre 96 gli stimoli rilevanti erano visivi.

I risultati del test mostrano un effetto congruenza generalmente più marcato per i soggetti sinesteti (maggiore velocità di risposta), anche se tale differenza risulta rilevante solo in riferimento alla classe di corrispondenze fono-simboliche, le quali, data la maggiore complessità degli stimoli, potrebbero essere considerate delle corrispondenze di più alto livello; per le altre due classi invece le performance dei due gruppi possono essere considerate simili.

Ad ogni modo, l'esito del test non ha portato a punti di svolta nel dibattito sulla sinestesia; rimane così invariata la versione di coloro che vedono i processi sinestetici come un qualcosa di totalmente slegato dai normali processi cognitivi cross-modali e quella di chi, invece, sostiene che entrambi gli elementi del puzzle siano semplicemente punti diversi di uno stesso continuum. Per i primi infatti, la differenza risiederebbe in una iperconnettività tra le aree corticali [33][150], e quindi la differenza tra sinesteti e non-sinesteti sarebbe strutturale e a livello neuronale; per i secondi, gli effetti di congruenza più accentuati, basati sulle migliori performance dei sinesteti nei test percettivi, ma comunque in molte circostanze equiparabili a quelli dei non-sinesteti, sarebbero la prova dell'esistenza di una base comune in entrambi i processi percettivi, semplicemente più forti e marcati nei soggetti sinestetici.

16 A titolo informativo, si ricorda che il cosiddetto *angolo di campo* costituisce il massimo angolo visivo che può coprire la vista, il quale mediamente oscilla tra un minimo di 40° e un massimo di 55°.

2.2 2017

Leonardo Bonetti, Marco Costa. *Pitch-verticality and pitch-size cross-modal interactions* [15]

Bonetti e Costa in due diversi studi hanno approfondito due coppie di corrispondenze crossmodali uditivo-visive, entrambe in rapporto all'altezza del suono: la prima tra pitch ed elevazione spaziale, la seconda tra pitch e dimensione delle forme.

Nel primo dei due test, trenta partecipanti (quindici di sesso maschile e quindici di sesso femminile, tutti senza alcun tipo di deficit uditivo e con un'età compresa tra i ventisei e i sessant'anni) furono sottoposti ad una serie di centosessanta prove (metà con stimoli congruenti e l'altra metà con stimoli incongruenti) accorpate in uno *speed test* volto a verificare la corrispondenza tra il pitch uditivo e la sua proiezione spaziale su un asse verticale. Sono stati impiegati solamente stimoli uditivi, che consistevano in segnali sinusoidali a diverse frequenze (100, 200, 600 e 800 Hz) tutti con pressione sonora fissa a 70 dBA. I vari segnali venivano riprodotti alternativamente su due altoparlanti, posizionati verticalmente in asse e rispettivamente a 60 cm sopra e sotto il livello dei padiglioni uditivi dei partecipanti, i quali sedevano a 50 cm dal fronte sonoro. Stando al modello derivato dai precedenti studi su questa classe di corrispondenza cross-modale, gli stimoli congruenti consistevano dunque nella riproduzione dei segnali a bassa frequenza (100 e 200 Hz) sull'altoparlante inferiore, mentre quelli ad alta frequenza (600 e 800 Hz) sull'altoparlante superiore. Ogni segnale acustico aveva durata pari a 1 s, distanziato dal successivo da una durata randomica compresa tra 1,5 e 3 s. Ai partecipanti, ad ogni presentazione dello stimolo uditivo, veniva chiesto di porre attenzione alla posizione del suono e di stabilire il più velocemente possibile se esso provenisse dal basso o dall'alto.

Il risultato mostra tempi di reazione più rapidi nell'identificazione degli stimoli congruenti, ma solamente per quelli ad alta frequenza, riprodotti quindi nell'altoparlante superiore. Questo è spiegabile dal semplice fatto che le onde sonore a bassa frequenza tendono ad espandersi in modo sferico (omnidirezionale), mentre all'aumentare della frequenza esse tendono ad assumere traiettorie più strette, viaggiando nel mezzo in modo via via più direzionale. Di conseguenza, nel test in questione, la particolare scelta a priori delle altezze dei segnali uditivi, non ha

sicuramente facilitato la performance dei partecipanti. Se sul piano orizzontale la posizione della sorgente sonora è facilmente discriminabile sulla base del ritardo interaurale con il quale il suono giunge a entrambi i padiglioni uditivi, sul piano verticale non vige lo stesso principio. Inoltre, questo test, a differenza di quelli precedenti, non ha utilizzato un secondo stimolo visivo proiettato contemporaneamente a quello uditivo, bensì è stato pianificato con una logica inedita. Probabilmente la compresenza di due stimoli, uditivi e visivi, facilita la valutazione poiché attiva in modo più diretto le funzioni della percezione cross-modale.

Nel secondo test coordinato da Bonetti e Costa furono coinvolti ventidue studenti universitari (metà di sesso maschile e metà di sesso femminile, tutti senza alcun tipo di deficit uditivo e con un'età compresa tra i ventiquattro e i sessant'anni). Gli stimoli uditivi consistevano in ventuno suoni sinusoidali le cui frequenze risultavano logaritmicamente spaziate tra loro (100, 125, 158, 199, 251, 316, 398, 501, 630, 794, 1000, 1258, 1584, 1995, 2511, 3162, 3981, 5011, 6309, 7943 e 10000 Hz), ciascuno della durata pari a 1 s. Questa volta i segnali acustici vennero fatti ascoltare in cuffia; ai partecipanti veniva chiesto di associare ogni stimolo uditivo ad una corrispettiva figura, che poteva essere scelta tra un set di nove cerchi di diverso diametro (0.41° , 0.66° , 1.33° , 2.45° , 4.69° , 6.92° , 8.73° , 11.51° e 14.32° , misure dell'angolo visuale) proiettati su uno schermo posto di fronte ai soggetti.

I risultati sono in linea con i risultati della letteratura precedente sul rapporto cross-modale tra pitch e dimensione delle forme. Nello specifico, la corrispondenza si è dimostrata valida su un range di frequenze molto esteso (100-10000 Hz), manifestandosi con una relazione del pitch inversamente proporzionale rispetto al diametro dei cerchi. Le condizioni di questo secondo test, stando ai risultati ottenuti, appaiono decisamente più favorevoli rispetto al primo; la presenza simultanea degli stimoli uditivi e visivi agevola la loro reciproca associazione su una base percettiva.

Irune Fernández-Prieto, Jordi Navarra. *The higher the pitch the larger its crossmodal influence on visuospatial processing* [37]

Come si è abbondantemente visto, diversi studi hanno confermato l’ipotesi che il pitch sia rappresentabile lungo un asse verticale: l’origine dei suoni acuti viene associata a posizioni spaziali più elevate rispetto a quella dei suoni gravi. Inoltre, da un punto di vista neurologico, i suoni acuti generano maggiori risposte neurali rispetto a quelli gravi, con conseguenti prestazioni migliori nella rilevazione di eventuali aumenti nell’altezza dei suoni. Ciò può facilmente condurre all’ipotesi che variazioni ascendenti del pitch uditivo abbia un impatto maggiore sull’*arousal* (attivazione) e/o sull’attenzione rispetto a stimoli uditivi con altezza discendente.

Fernández-Prieto e Navarra hanno analizzato se queste differenze neurali hanno implicazioni, a livello cognitivo, nella rappresentazione spaziale dell’altezza del suono; in particolare, l’ipotesi oggetto di indagine riguardava la possibilità che stimoli uditivi con altezze ascendenti potessero essere maggiormente interpretabili per la creazione di rappresentazioni spaziali rispetto a segnali acustici ad altezza discendente.

Nella realizzazione dello studio gli autori hanno utilizzato una versione modificata del paradigma di Posner, originariamente pensato per valutare l’attenzione selettiva spaziale nella sola modalità visiva¹⁷. La modifica del paradigma creata da Fernández-Prieto e Navarra prevedeva l’utilizzo di stimoli uditivi ascendenti (200-700 Hz) o discendenti (700-200 Hz), con il presupposto che avessero una diversa capacità nel modulare lo stato di attivazione (*arousal*) e vigilanza dei soggetti e la loro attenzione visuo-spaziale. Nello specifico, l’ipotesi alla base dello studio era la seguente: la concorrente rilevazione di stimoli visivi sarebbe maggiormente influenzata da suoni ascendenti piuttosto che da suoni discendenti.

17 Nella versione ideata da Posner venivano presentati due quadrati, sulla sinistra e sulla destra di uno schermo, uno dei quali era evidenziato (segnaletica spaziale); dopo un certo ritardo, nello stesso quadrato precedentemente evidenziato (prova congruente) o nel quadrato opposto (prova incongruente), compariva un asterisco (stimolo visivo). I soggetti si dimostrarono più rapidi nell’individuazione dell’asterisco nelle prove congruenti, stabilendo che un precedente segnale visuo-spaziale che orienti l’attenzione selettiva può facilitare l’individuazione di uno stimolo visivo successivo. Per approfondire si veda l’articolo dell’autore riportato in bibliografia [108].

A sedici soggetti normo-udenti e normo-vedenti e senza alcuna preparazione musicale, è stato chiesto di individuare uno stimolo visivo (un asterisco bianco) proiettato in un monitor il più velocemente e accuratamente possibile. Lo stimolo visivo poteva apparire indifferentemente in quattro posizioni diverse: in alto a sinistra, in alto a destra, in basso a sinistra e in basso a destra. La comparsa dell'asterisco veniva preceduta dalla presentazione di suoni ascendenti o discendenti: gli stimoli visivi ed uditivi erano distanziati l'uno dall'altro in modo variabile (400 o 550 ms) al fine di impedire una certa prevedibilità nella loro apparizione, e il test comprendeva 320 prove.

I risultati ottenuti nello studio suggeriscono che, in alcune condizioni, la rilevazione spaziale degli stimoli visivi viene influenzata dalle proprietà degli stimoli uditivi. In particolare, i segnali acustici ascendenti hanno originato risposte più rapide per gli stimoli visivi presentati nella parte superiore dello schermo rispetto a quelli presentati nella parte inferiore; tale modulazione dei tempi di risposta che intercorre tra stimoli uditivi ascendenti e discendenti confermerebbe dunque che le corrispondenze crossmodali uditivo-visive con suoni ascendenti hanno una maggiore capacità di attivare l'attenzione e sono in grado generare delle rappresentazioni spaziali più nitide.

Infine, non è stato rilevato alcun andamento particolare per quanto concerne la discriminazione degli stimoli tra lato sinistro e destro, e ciò supporterebbe una considerazione d'ordine generale, cioè che il pitch sia meglio rappresentato da un'elevazione spaziale piuttosto che una sua dislocazione su un asse orizzontale.

Alessia Tonelli, Luigi F. Cuturi, Monica Gori. *The Influence of Auditory Information on Visual Size Adaptation* [136]

La percezione delle dimensioni degli stimoli visivi può essere influenzata da molteplici fattori. Ad esempio, la *size adaptation* descrive la tendenza del sistema visivo a regolare la reattività neurale nelle dimensioni delle rappresentazioni dopo un'esposizione prolungata a particolari stimoli visivi. Una presentazione prolungata di uno stimolo visivo (chiamato *adapter*) con dimensioni costanti è infatti in grado di influenzare la dimensione percepita di uno stimolo visivo mostrato successivamente (*reference*) [151]; se l'*adapter* è più grande del *reference*, la dimensione percepita di quest'ultimo risulterà più piccola della sua reale dimensione fisica, viceversa apparirà più grande nel caso in cui lo stimolo iniziale sia più piccolo. Tonelli, Cuturi e Gori, attraverso il test qui riportato, hanno sperimentato la possibilità di modulare gli effetti visivi della *size adaptation* attraverso determinati stimoli visivi e uditivi concorrenti.

Nel test sono stati coinvolti sedici adulti di ambo i sessi (età media di venticinque anni). Gli stimoli visivi ai quali sono stati sottoposti consistevano in cerchi di varie dimensioni (etichettati come *adapter*, *test* e *reference*) proiettati su un monitor posizionato a 57 cm di distanza dal punto di osservazione. Gli stimoli uditivi erano due suoni sinusoidali, uno a 100 e l'altro a 9000 Hz di frequenza, entrambi di durata pari a 500 ms. Il test si sviluppava in due fasi, ognuna delle quali ulteriormente suddivisa in due blocchi di prove, distinti dalla presenza o meno degli stimoli uditivi. Nella prima fase ai partecipanti venivano mostrati due cerchi, chiamati rispettivamente *test* e *reference*, ognuno per la durata di 500 ms. Il diametro del primo cerchio (*test*) poteva variare da un minimo di 0.5 ad un massimo di 1.5 volte quello del secondo cerchio (*reference*). Ai soggetti, per ogni prova veniva chiesto di indicare se il primo stimolo visivo fosse percepito con dimensioni maggiori o minori rispetto al secondo. Il secondo blocco di prove di questa prima fase prevedeva la presenza degli stimoli uditivi, presentati in modo sincrono assieme al primo cerchio.

La seconda fase del test era uguale rispetto alla prima, salvo per la presentazione di un terzo stimolo visivo (*adapter*, con diametro maggiore rispetto agli altri due) mostrato per 10 s prima degli altri due (*test* e *reference*).

Il test mirava dunque a comprendere le modulazioni apportate dalla stimolazione uditiva negli effetti visivi di *size adaptation*, mettendo a confronto i risultati delle prove in presenza e in assenza di segnali acustici concorrenti.

In entrambe le fasi del test (con e senza stimolo adattivo iniziale), il confronto con i risultati ottenuti in assenza di stimolazione uditiva stabilisce che la presenza del segnale acustico modifica la percezione della reale dimensione dello stimolo visivo simultaneo.

Nella prima fase, caratterizzata dall'assenza di stimoli visivi in grado di procurare effetti di *size adaptation*, la presenza del segnale acustico a bassa frequenza faceva apparire l'immagine concorrente con dimensioni maggiori rispetto alle stesse prove senza stimoli uditivi; gli stimoli ad alta frequenza, invece, rendevano le dimensioni percepite del primo cerchio più piccole.

Nella seconda fase, gli effetti di *size adaptation* venivano altresì alterati dalla presenza degli stimoli uditivi. Nella condizione in cui lo stimolo adattivo (*adapter*) di grandi dimensioni veniva seguito dal segnale ad alta frequenza, lo stimolo visivo concorrente (*test*) veniva percepito con dimensioni minori rispetto alle prove con soli stimoli visivi; nella condizione contraria, cioè quando lo stimolo adattivo di piccole dimensioni veniva seguito dal frammento sonoro a bassa frequenza, l'immagine *test* veniva percepita con dimensioni maggiori rispetto alle medesime prove nelle condizioni esclusivamente visive.

Questo sta a significare che la compresenza di stimoli visivi e uditivi è in grado di ridurre gli artefatti visivi di *size adaptation*. In generale, ciò dimostrerebbe che il pitch ha un effetto cross-modale non solo nella discriminazione dimensionale degli stimoli visivi percepiti in modo sincrono, ma anche su quelli percepiti in seguito ai processi di adattamento visivo. Tali funzioni cognitive spiegherebbero come il suono sia in grado di modulare le funzioni della corteccia visiva primaria (V1) nelle prime fasi di elaborazione delle immagini.

Y. Jamal, S. Lacey, L. Nygaard, S. Sathian. *Interactions Between Auditory Elevation, Auditory Pitch and Visual Elevation During Multisensory Perception* [64]

Come reca il titolo dell'articolo scritto a quattro mani da Yaseen Jamal, Simon Lacey, Lynne Nygaard e Krish Sathian, il test realizzato dagli stessi autori ha cercato di far luce sul mutuo rapporto fra tre diverse *features* uditive e visive: elevazione uditiva (AE, *auditory elevation*), pitch uditivo (AP, *auditory pitch*) ed elevazione visiva (VE, *visual elevation*).

Nel test sono stati coinvolti ventiquattro partecipanti, dodici di sesso maschile e dodici di sesso femminile, aventi un'età media di ventisei anni e senza alcuna disfunzione uditiva e visiva. Gli stimoli visivi erano due cerchi bianchi (RGB = 240, 240, 240) aventi diametro pari a 4 cm e proiettati su sfondo nero, collocati rispettivamente al margine superiore e inferiore di un monitor LCD. Gli stimoli uditivi impiegati nel test erano due, entrambi della durata pari a 500 ms e ottenuti da un'operazione di filtraggio del rumore bianco: un segnale ottenuto con un filtro passa alto con frequenza di taglio a 8 kHz, l'altro ottenuto con un filtro passa basso con frequenza di taglio a 1 kHz; essi erano riprodotti alternativamente da due speaker posti sopra e sotto il monitor.

Per poter studiare la reciproca influenza delle tre *features* audio-visive, gli stimoli sono stati organizzati in modo tale da creare otto diverse combinazioni, prevedendo condizioni sia congruenti sia incongruenti (vedi tabella numero 4). Tenendo monitorato il tempo di reazione dei partecipanti, veniva chiesto loro di classificare nel modo più rapido e preciso possibile la posizione della *feature* rilevante tra due possibilità (alto o basso), ignorando le due *features* complementari, su un totale di sessantaquattro prove. Giusto per fare un esempio, in una prova era chiesto di classificare come alto o basso lo stimolo visivo (il cerchio proiettato in alto o in basso) mentre lo stimolo uditivo (alto o basso) veniva riprodotto su uno dei due speaker (rispettivamente sopra o sotto il monitor). In questo esempio, la *feature* rilevante era l'elevazione visiva, mentre le *features* non rilevanti e concomitanti erano il pitch e l'elevazione uditiva.

| <i>Congruenze</i> | <i>Pitch uditivo</i> | <i>Elevazione uditiva</i> | <i>Elevazione visiva</i> |
|------------------------|----------------------|---------------------------|--------------------------|
| AP-AEc, AP-VEc, AE-VEc | alto | alto | alto |
| | basso | basso | basso |
| AP-AEc, AP-VEi, AE-VEi | alto | alto | basso |
| | basso | basso | alto |
| AP-AEi, AP-VEc, AE-VEi | alto | basso | alto |
| | basso | alto | basso |
| AP-AEi, AP-VEi, AE-VEc | basso | alto | alto |
| | alto | basso | basso |

Table 4: Possibili combinazioni di stimoli congruenti (c) e incongruenti (i) predisposte nel test di Jamal, Lacey, Nygaard e Sathian.

Prima del test ufficiale, per controllare la capacità di discriminazione delle singole *features* dei soggetti (pitch, elevazione uditiva e elevazione visiva), i partecipanti sono stati sottoposti a stimoli unimodali ed era loro chiesto di classificarli come alti o bassi. Dai risultati si evince che tutte e tre le *features*, se presentate da sole, sono percepite correttamente in ugual misura.

Il test vero e proprio cercava poi di evidenziare se l'incongruenza di stimoli concorrenti generasse effetti di interferenza o meno. I risultati mettono in luce un primo dato saliente: quando viene chiesto di classificare l'elevazione spaziale dello stimolo visivo, stimoli simultanei incongruenti non pregiudicano la corretta discriminazione di tale *feature*, che avviene inoltre in modo più veloce e preciso rispetto alle altre due *features*; ciò confermerebbe il ruolo dominante che il canale visivo ricalca nell'attività percettiva dell'uomo, così come viene dimostrato anche dal cosiddetto "effetto ventriloquo". Un secondo dato, abbastanza scontato, deriva dal confronto dei tempi di reazione tra stimoli congruenti e incongruenti: la combinazione congruente delle tre *features* audio-visive velocizza in modo consistente la risposta dei soggetti rispetto alla combinazione di stimoli incongruenti, i quali darebbero infatti luogo a effetti di interferenza. Infine, l'ultimo risultato saliente dice che un'incongruenza intra-modale tra il pitch e l'elevazione uditiva è in grado di iniziare la corretta interpretazione della corrispondenza cross-modale tra pitch uditivo ed elevazione visiva; tale risultato, comunque, risulta consistente solo quando l'attenzione dei partecipanti è diretta nei confronti delle *features* sonore (AP, AE), accentuata quando la dimensione rilevante dello stimolo uditivo è rappresentata dal pitch, mentre rimane invariata la corretta interpretazione della posizione spaziale delle forme anche quando gli stimoli concorrenti (AP, AE) siano reciprocamente incongruenti.

E. Bruzzi, F. Talamini, K. Priftis, M. Grassi. A SMARC Effect for Loudness [22]

Gli autori di questo test hanno approfondito l'effetto SMARC (*Spatial-Musical Association of Response Codes*) in rapporto alla loudness uditiva. Nei precedenti paragrafi è stato evidenziato come tale corrispondenza spaziale uditivo-visiva sia stata lungamente studiata in rapporto al pitch; in questo caso, invece, si è cercato di indagare una possibile corrispondenza tra la loudness e l'elevazione spaziale. Precedentemente la corrispondenza tra loudness e posizione spaziale è stata studiata in sole tre occasioni, due delle quali misero a confronto tale proprietà uditiva con una corrispondente proiezione su un asse orizzontale [26][60], mentre nella terza sottoponendola ad un confronto con l'elevazione spaziale [38]. Da tali studi appare congruente l'associazione tra la loudness e la posizione orizzontale in modo direttamente proporzionale tra le due *features* con un andamento da sinistra a destra (suoni deboli a sinistra, suoni forti a destra), mentre un'altrettanta congruenza emerge tra la loudness e la proiezione verticale in modo direttamente proporzionale dal basso verso l'alto (suoni deboli in basso, suoni forti in alto).

Bruzzi, Talamini, Priftis e Grassi (2017) hanno coordinato un ulteriore studio per approfondire la corrispondenza tra la loudness e l'elevazione visiva. La ricerca sperimentale coinvolse quarantaquattro studenti normo-udenti dell'Università di Padova, nessuno dei quali in possesso di competenze musicali. Gli stimoli erano solo uditivi e consistevano in un segnale di riferimento e di altri sei segnali di controllo. Lo stimolo di riferimento era un suono complesso della durata di 150 ms, con ampiezza pari a 70 dB SPL, con frequenza fondamentale a 550 Hz e uno spettro composto da cinque armoniche superiori. Gli altri stimoli di controllo erano copie del segnale di riferimento ma con ampiezza variata: tre con ampiezza minore (-6, -12 e -18 dB) e tre con ampiezza maggiore (+6, +12 e +18 dB).

I soggetti venivano bendati e i vari segnali venivano fatti loro ascoltare attraverso cuffie di alta qualità. Ad ogni prova (sessanta in tutto), il segnale di riferimento precedeva quelli di controllo, separati da un silenzio di 500 ms; ad ogni coppia di stimoli ai partecipanti veniva chiesto di indicare nel modo più veloce e preciso possibile se il secondo suono fosse percepito come più forte o più debole rispetto al primo. Per comunicare tale dato essi dovevano premere due pulsanti disposti uno sopra l'altro in una periferica costruita ad hoc per l'occasione. Il test è stato pensato in due sessioni separate, in modo da poter sperimentare una condizione

considerata “compatibile” e una considerata invece “incompatibile”: nella condizione compatibile (o congruente) se il secondo stimolo di controllo veniva percepito più forte rispetto al primo i partecipanti dovevano premere il bottone superiore e se invece veniva percepito come più debole dovevano premere quello inferiore; nella condizione incompatibile (o incongruente) se il secondo stimolo veniva percepito più forte i partecipanti dovevano segnalarlo premendo il tasto inferiore, mentre se il suono appariva più debole il tasto da premere era quello superiore. Dunque, la condizione di compatibilità o incompatibilità è stata pensata sulla base delle precedenti ricerche, che, come è stato evidenziato, stabilivano la congruenza tra suoni ad alta intensità con una loro proiezione verso l’alto, e viceversa per i suoni a bassa intensità una loro collocazione verso il basso.

I risultati hanno evidenziato risposte più rapide nella condizione compatibile, ancor più rapide con differenze di intensità maggiori. Ciò confermerebbe l’esistenza di un effetto SMARC anche in rapporto alla loudness, non solo, dunque, legata al pitch. Tale corrispondenza potrebbe essere causata da un’implicita connessione tra le due *features* in riferimento alla loro rispettiva magnitudine: “più alto” significherebbe dunque “più forte”. Da non sottovalutare una possibile derivazione linguistica di tale effetto cross-modale; è uso comune definire “ad alto volume” un suono ad alta intensità, mentre un suono debole viene definito “a basso volume”.

Nan Shang, Suzy J. Styles. *Is a High Tone Pointy? Speakers of Different Languages Match Mandarin Chinese Tones to Visual Shapes Differently*
[120]

Nan Shang e Suzy Styles hanno recentemente implementato un duplice test online per verificare la corrispondenza tra le proprietà uditive e visive mediante una rivisitazione del classico effetto *bouba/kiki* su tre diversi gruppi linguistici, uno cinese, uno inglese e uno bilingue anglo-cinese, servendosi dell’inflessione dell’altezza dei quattro toni tipici del mandarino cinese. Brevemente, la lingua cinese è una lingua a toni, cioè una lingua che prevede la diversa significazione di una stessa parola sulla base di una diversa inflessione del tono lessicale, che può essere di quattro tipi, chiamati rispettivamente tono 1, 2, 3 e 4¹⁸: il Tono 1 è acuto e ad altezza costante, il Tono 2 ha un andamento ad altezza crescente, il Tono 3 ha un andamento misto, discendente e poi ascendente, e infine il Tono 4 si caratterizza per un’altezza discendente.

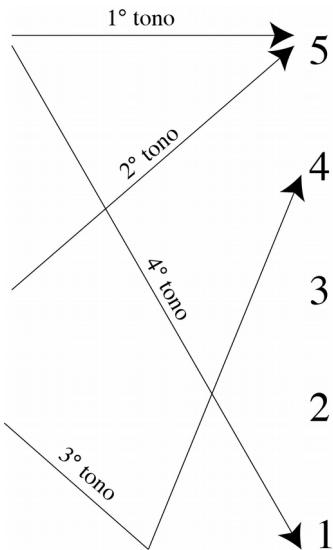


Illustration 19: Inflessioni dei quattro toni cinesi.

Gli autori del test in questione hanno utilizzato come stimoli uditi i campionamenti delle vocali /i/ e /u/, pronunciate da una madrelingua cinese di sesso femminile,

¹⁸ Ad esempio, la parola *yan*, a seconda del tono con il quale viene pronunciata, può voler significare: fumo (*yan1*), sale (*yan2*), occhio (*yan3*), colorato (*yan4*).

entrambe nelle loro quattro varianti tonali, per un totale di otto diversi stimoli uditivi. La tabella numero 5 mostra la misurazione delle rispettive altezze.

| Stimolo | Pitch onset (Hz) | Pitch offset (Hz) | Mean pitch (Hz) |
|---------|------------------|-------------------|-----------------|
| u1 | 281.0 | 301.0 | 286.4 |
| u2 | 236.4 | 294.2 | 229.5 |
| u3 | 209.3 | 213.3 | 166.1 |
| u4 | 308.1 | 208.8 | 220.6 |
| i1 | 279.7 | 307.7 | 296.9 |
| i2 | 223.1 | 290.4 | 229.3 |
| i3 | 202.2 | 189.4 | 181.9 |
| i4 | 280.5 | 212.3 | 224.5 |

Table 5: Altezze iniziali, finali e medie degli stimoli uditivi del test di Shang e Styles.

Come stimoli visivi sono stati preparati due diversi solidi tridimensionali sulla base dei tratti fondamentali del paradigma *bouba/kiki*, una simil-sfera dalla geometria irregolare e un solido dalla superficie tentacolare e acuminata¹⁹.



Illustration 20: Stimoli visivi impiegati nel test.

© Shang & Styles.

Nel primo dei due studi sono stati coinvolti cento partecipanti, quarantacinque di lingua cinese, trenta bilingue inglese-cinese e venticinque inglesi. Senza alcun limite di tempo, i partecipanti ascoltavano in cuffia una ad una le otto varianti vocaliche, ciascuna presentata assieme a entrambe le figure; veniva dunque chiesto loro di indicare quale delle due consideravano più coerente rispetto allo stimolo uditorio.

19 Stimoli visivi e uditivi sono liberamente consultabili e scaricabili al seguente url:
<https://osf.io/364fm/>

I risultati di questo primo test confermano le corrispondenze cross-modali tra il pitch e la geometria delle forme stabilite dai precedenti test percettivi: tutti i differenti gruppi linguistici hanno associato le diverse inflessioni tonali delle /i/ alla figura più regolare, mentre le quattro diverse /u/ alla figura più irregolare. Dunque è stata confermata la corrispondenza suoni acuti-geometrie appuntite e suoni gravi-geometrie arrotondate anche tra soggetti appartenenti a diversi ceppi linguistici.

Nel secondo test sono stati utilizzati gli stessi stimoli uditivi e visivi; questa volta però gli stimoli visivi venivano presentati uno alla volta, ognuno accompagnato da una coppia di stimoli uditivi distanziati tra loro da una breve pausa di 500 ms. Ai toni 1 venivano contrapposti i toni 3 e 4, ottenendo così delle coppie di stimoli uditivi in cui ad un suono vocalico ad altezza costante veniva contrapposto un suono vocalico ad altezza variata. I partecipanti erano diversi rispetto al primo test, ma, in modo analogo, appartenenti ai tre diversi ceppi linguistici (trentuno di lingua cinese, ventiquattro bilingue e quarantanove di lingua inglese). Ad ogni figura veniva chiesto ai partecipanti di indicare il suono più congruente tra la coppia ad essa associata, anche in questo caso senza alcun limite di tempo.

Diversamente dal primo test, il secondo ha evidenziato una sostanziale contrapposizione tra il gruppo madrelingua cinese e quello inglese. I soggetti appartenenti a quest'ultimo gruppo mostrano delle risposte coerenti con il modello verificato nel primo test, mantenendo inalterata l'associazione tra la figura appuntita e i suoni vocalici più acuti (tono 1) e la forma sferica a quelli con frequenza media più bassa (toni 3 e 4). Il gruppo di lingua cinese invece mostra un comportamento insolito: esso infatti sembrerebbe molto più attento alla variazione dell'altezza dei singoli suoni, attribuendo ai toni 1, statici in altezza, la figura sferica, mentre i toni 3 e 4 con altezza modulata vengono associati alla figura più irregolare, cioè quella appuntita.

La differenza emersa dal secondo test potrebbe essere spiegata dal fatto che, per sua natura, la lingua cinese richiede un'accurata attenzione alle inflessioni di tono, poiché esse determinano significati diversi per una stessa parola. Quindi i soggetti di lingua cinese sembrerebbero prestare più attenzione e creare dei paralleli tra il profilo dell'altezza e la regolarità/irregolarità degli stimoli visivi.

Infine, tale differenza tra differenti gruppi linguistici risulta antitetica rispetto alla posizione dominante nel campo del simbolismo fonetico, la quale sostiene l'universalità delle associazioni fono-simboliche. Ceppi linguistici con differenti caratteristiche prosodiche producono diverse corrispondenze simboliche.

G. Hamilton-Fletcher, C. Witzel, D. Reby, Jamie Ward. *Sound Properties Associated With Equiluminant Colours* [57]

Nel test sperimentale qui riportato gli autori hanno investigato il rapporto assai controverso tra suono e colore; riallacciandosi ai risultati di precedenti test su tale tema, l'intento della ricerca era quello di fare maggior chiarezza sui parametri del colore meno presi in considerazione: tonalità e saturazione. Se test pregressi avevano infatti compreso come il pitch e la loudness venissero generalmente correlati al livello di luminosità visiva (brillanza), tonalità e saturazione erano spiegati da risultati incerti. Le ipotesi iniziali degli autori erano le seguenti: un aumento nel pitch uditivo sarebbe stato associato ad un aumento nel livello di saturazione dei colori, con una tendenza verso le tonalità vicine al giallo, tendenza che si sarebbe dovuta verificare in modo simile entro diverse magnitudini del range frequenziale; inoltre, basandosi sui risultati ottenuti dal test di Giannakis e Smith discusso nel precedente capitolo [52], gli autori ipotizzavano di confermare una corrispondenza direttamente proporzionale tra il livello di saturazione e la loudness. Infine, sottponendo i soggetti a diversi timbri sonori, gli autori prevedevano di riscontrare parallele variazioni nella tonalità cromatica e nella saturazione, differenze determinate da diversi “centri di gravità” dello spettro sonoro (ovverosia diversi *centroidi spettrali* [138]): stimoli uditivi con uno spettro più carico di alte frequenze dovrebbero essere, secondo gli autori, associati alle tonalità più sature e vicine al giallo rispetto a suoni con uguale altezza ma con timbri più poveri di armoniche superiori, la cui cromia corrispondente dovrebbe tendere al blu.

Al test hanno partecipato quarantaquattro studenti della University of Sussex, tutti senza alcuna deficienza visiva e uditiva, nessuno dei quali aveva mai manifestato comportamenti percettivi sinestetici.

Gli stimoli uditivi sono stati preparati secondo diversi criteri morfologici. Una prima categoria consisteva di suoni sinusoidali con loudness costante; essi sono stati divisi in due gruppi di altezze diverse, in modo tale da poter verificare le reazioni dei partecipanti su range frequenziali più o meno ampi. Un primo gruppo copriva un range da 100 a 3200 Hz (100, 200, 400, 800, 1600 e 3200 Hz); il secondo gruppo copriva un range ridotto, da 440 a 880 Hz (440, 493, 523, 587, 659, 698, 784 e 880 Hz). Una seconda categoria di stimoli uditivi era stata predisposta per studiare le *features* del colore in relazione alla variazione di loudness; partendo da un suono

sinusoidale a 400 Hz e con loudness pari a 40 phon, sono stati generati tre ulteriori segnali con frequenza invariata ma con loudness avente un rapporto rispettivamente pari a metà (0.5), un quarto (0.25) e un decimo (0.1) rispetto ai 40 phon del segnale originale. La terza categoria di stimoli uditivi era volta a comprendere gli effetti cross-modali tra timbro e colore; gli autori hanno dunque sintetizzato un suono vocalico maschile con frequenza fondamentale a 83 Hz, successivamente sottoposto a cinque diverse operazioni di filtraggio spettrale con un filtro passa-banda settato in altrettante configurazioni: 100-200 Hz, 200-400 Hz, 400-800 Hz, 800-1600 Hz, 1600-3200 Hz. Infine, un'ultima categoria di segnali uditivi è stata preparata allo scopo di comprendere l'influenza del centro di gravità spettrale sulla scelta dei colori; un primo suono è stato generato con sintesi additiva, sommando assieme sei componenti sinusoidali (100, 200, 400, 800, 1600 e 3200 Hz) e ulteriori cinque segnali sono stati derivati da questo alterando di volta in volta l'ampiezza delle singole componenti armoniche. In modo analogo, è stato riutilizzato il suono vocalico precedentemente creato: ne sono state prodotte cinque copie, ognuna delle quali con una diversa equalizzazione spettrale.

Ad ogni prova un unico stimolo visivo veniva riprodotto da un monitor posto di fronte ai partecipanti; esso consisteva in un cerchio colorato mantenuto con brillanza costante con una dimensione pari a 2.55° di angolo visuale e proiettato su sfondo grigio. La definizione della tonalità e della saturazione era autonomamente gestita dagli stessi partecipanti, i quali avevano di volta in volta il compito di settare tali parametri nel modo ritenuto di volta in volta più opportuno, sulla base degli stimoli uditivi concorrenti. Ogni partecipante, in diverse prove, veniva sottoposto a tutte le diverse famiglie di segnali acustici.

Per quanto concerne gli effetti del pitch sul livello di saturazione, i risultati confermano l'ipotesi degli autori: un aumento del pitch corrisponde ad un aumento di saturazione; i grafici dei risultati mostrano una corrispondenza con andamento pressoché logaritmico nel set di segnali sinusoidali con range 100-3200 Hz, mentre assume un andamento proporzionalmente lineare nel set con range 440-880 Hz. Il fatto che per entrambi i gruppi di segnali uditivi venga mantenuto un andamento generale simile nella corrispondenza tra pitch e saturazione, farebbe pensare che la mappatura di tali *features*, sebbene la regola generale rimanga invariata, sia comunque relativa e scalabile in base al contesto uditivo: un allargamento o un restringimento dei margini frequenziali, infatti, sposta di conseguenza la mappatura del livello di saturazione.

Anche i risultati relativi alla loudness mostrano un andamento direttamente proporzionale tra le due *features*; livelli di loudness più elevati sono considerati congruenti ai colori più saturi. Le previsioni degli autori non sono state tradite anche dai risultati sul baricentro spettrale dei timbri complessi; a parità di altezza, più il centroide si sposta verso il registro acuto, più la saturazione viene impostata a livelli elevati. Tuttavia, mettendo a confronto le risposte ottenute trasversalmente dalle diverse categorie di stimoli uditivi, la saturazione appare meglio rappresentabile attraverso i segnali sinusoidali, mentre il baricentro spettrale ha una minore capacità di attivare la corrispondenza cross-modale in relazione alla saturazione del colore.

Passando ai risultati ottenuti dallo studio sulla tonalità, la variazione cromatica appare sensibile alle variazioni del pitch: suoni a bassa frequenza sono statisticamente rappresentati dalle tonalità cromatiche vicine al viola-blu, mentre all'aumentare del pitch vi è una progressiva variazione verso il giallo, tonalità complementare al viola-blu. Tuttavia questo dato non può essere considerato come realmente oggettivo, poiché alcuni partecipanti hanno dato risposte in netto contrasto rispetto a tale trend generale; vi è dunque, per quanto concerne il rapporto tra pitch e tonalità, un certo margine di soggettività, che non permette di stabilire una regola percettiva generale e per tutti valida.

Nei risultati delle prove con stimoli uditivi variabili nella loudness, all'aumentare dell'intensità sonora percepita si nota una progressiva e generale variazione cromatica lungo un asse blu-giallo. Dati meno salienti non emergono invece dalle prove con stimoli uditivi ad altezza fissa ma con baricentro spettrale variato.

Questo test in primo luogo ha dimostrato per la prima volta che, mantenendo il livello della brillanza costante, le *features* uditive possono trovare una corrispondenza tra le altre dimensioni percettive del colore, la tonalità e la saturazione. Variazioni di loudness, pitch e centroide spettrale, danno luogo a più corrispondenze, le quali necessitano tuttavia di ulteriori approfondimenti. Un dato sicuramente importante emerge dalla relatività e scalabilità degli intervalli di valori di tali *features*, in particolar modo in riferimento al pitch; i soggetti si sono mostrati in grado di adattare un modello di base (suoni gravi-blu, suoni acuti-giallo) a seconda del set di stimoli ai quali di volta in volta venivano sottoposti. Da ciò è possibile ipotizzare che le associazioni cross-modali di questo tipo siano basate sulla polarizzazione di due estremi, i cui valori sono scalabili e adattabili a più contesti.

Clare Jonas, Mary Jane Spiller, Paul Hibbard. *Summation of visual attributes in auditory-visual crossmodal correspondences* [65]

Jonas, Spiller e Hibbard prendono le mosse dalla speculazione secondo la quale praticamente tutti i test percettivi indirizzati allo studio della percezione cross-modale definiscono a priori e utilizzino due sole *features*; questo sarebbe, a buon dire da parte degli autori, una semplificazione impossibile in natura, dove invece gli eventi sono decisamente più complessi e multidimensionali.

Si sono allora chiesti come interagiscono fra loro più classi di corrispondenze cross-modali e, soprattutto, se vi sia un principale modello che spieghi tale complessità nell'interazione sensoriale. Per il test sono quindi state selezionate quattro *features* visive: brillanza, saturazione, dimensione ed elevazione spaziale, tutte messe in relazione al pitch uditivo. Gli stimoli visivi consistevano in cerchi colorati di quattro tonalità fisse, indicate con sistema HSL²⁰ (rosso 0°, giallo 58°, verde 120° e blu 240°) proiettati su sfondo grigio; le *features* da studiare invece potevano variare su un range di tre valori (basso, medio, alto), rispettivamente: luminosità e saturazione (0.16, 0.5 e 0.85), dimensione (\varnothing 8, 16 e 25 mm) ed elevazione spaziale, calcolata come distanza rispetto il margine superiore dello schermo di proiezione (56, 40 e 23 mm). Gli stimoli uditivi preposti erano solamente due, entrambi di 1 s di durata ma con due altezze diverse: 261.63 Hz e 523.25 Hz.

Il test è stato predisposto su una piattaforma online, dove sono stati coinvolti centotredici partecipanti di età compresa tra i diciotto e i sessantasette anni. Ogni soggetto ha utilizzato un dispositivo elettronico di sua proprietà (computer, tablet o smartphone), nelle condizioni che riteneva più opportune.

I partecipanti venivano sottoposti ad un totale di ottanta prove, suddivise su quattro differenti livelli. In ogni prova venivano mostrati per primi gli stimoli visivi, due cerchi con valori diversi delle stesse *features*; i soggetti potevano poi riprodurre i due stimoli uditivi il numero di volte ritenuto opportuno, senza alcun limite di tempo. Per ogni prova veniva chiesto di associare ogni suono ad una rispettiva figura.

²⁰ Il sistema HSL (acronimo di *Hue* [tonalità], *Saturation* [saturazione], *Lightness* [luminosità]) utilizza una misura angolare per la tonalità (su cerchio cromatico) e una scala da 0.0 a 1.0 per saturazione e luminosità, valori comodamente convertibili in dato percentuale.

Nelle prove relative al primo livello le *features* in gioco erano due, il pitch e una delle quattro *features* visive. Ad ogni livello successivo aumentava la complessità, con l'aggiunta di una *feature* visiva; in questo modo, all'ultimo stadio del test tutte e cinque le *features* erano presentate assieme.

I risultati del test sono in linea con i principali modelli di base delle varie corrispondenze cross-modali quando una sola *feature* visiva viene contrapposta al pitch uditivo: il segnale con frequenza maggiore è stato infatti associato a stimoli visivi con alta luminosità, alta saturazione, piccole dimensioni e spazialmente elevato. Con la compresenza di tutte e quattro le *features* visive, viene confermata l'ipotesi degli autori secondo cui il sistema percettivo opererebbe una somma delle *features* disponibili per una ricostruzione unitaria dell'evento multidimensionale (*cross-modal binding*), e non vi sia una gerarchizzazione degli stimoli visivi per la quale particolari *features* vengano considerate più importanti di altre. Tuttavia si evince un'interferenza generata dalla saturazione sulla corretta valutazione della luminosità nei suoi valori di picco massimo e minimo; questo dato dimostrerebbe ancora una volta come le *features* degli stimoli colorati vengano percepite come integrali, dunque difficilmente separabili tra loro.

In ultima istanza è doveroso fare un appunto sulla metodologia adottata. Il metodo “fai da te” online che sfrutta la tecnologia mobile in possesso dalla maggioranza della popolazione, se da una parte permette di coinvolgere un maggior numero di partecipanti, dall'altra non può garantire condizioni ottimali e controllate durante lo svolgimento delle prove; un test che coinvolge una tale varietà di *features* percettive come quello in questione non dovrebbe essere svolto con tali modalità, poiché il rischio di ottenere dati corrotti e falsati da condizioni non consone è davvero troppo alto.

2.3 2018

S. C. Kwok, C. Fantoni, L. Tamburini, L. Wang, W. Gerbino. A biphasic effect of cross-modal priming on visual shape recognition [71]

Kwok, Fantoni, Tamburini, Wang e Gerbino hanno testato gli effetti cross-modali sulla memoria visiva a breve termine; nello specifico, essi si sono posti l'obbiettivo di comprendere quale ruolo giochi la percezione cross-modale di stimoli audio-visivi sulla cosiddetta *inibizione di ritorno* (*Inhibition of Return*, IOR): brevemente, in psicologia la IOR si riferisce all'attenzione selettiva rivolta agli oggetti percepiti dalla visione periferica, la quale può essere correttamente mantenuta fino ad un massimo di cinque elementi visivi²¹ in rapida successione tra loro; oltre tale numero, la percezione di ulteriori stimoli visivi tende ad inibire la memoria nei confronti degli stimoli visivi precedentemente individuati a livello spaziale. Gli autori del test in questione sono partiti dal presupposto che la corrispondenza cross-modale tra il timbro e la geometria delle forme possa influenzare l'elaborazione di una sequenza di stimoli visivi.

Al test hanno partecipato diciotto soggetti, otto di sesso femminile e dieci di sesso maschile, aventi un'età media di quarantanove anni, tutti senza alcuna disfunzione né uditiva né visiva.

Gli stimoli visivi impiegati nell'esperimento consistevano in venti figure di color grigio chiaro (20.7 cd/m^2) proiettate su uno schermo con sfondo scuro (5 cd/m^2), progettate secondo il paradigma *bouba/kiki*, dieci delle quali con profilo appuntito, mentre le altre dieci con un profilo arrotondato. Un ulteriore stimolo visivo aveva una funzione di disturbo e di mascheramento, e consisteva in un'immagine rumorosa della stessa dimensione delle figure appuntite e arrotondate, composta da oltre duecento piccoli quadrati di diversi grigi. Gli stimoli uditivi comprendevano due suoni, ognuno di essi pensato per essere congruente ad una corrispettiva categoria di figure: un suono etichettato come “tzk” (previsto per essere congruente alle immagini spigolose), rumoroso e composto da improvvisi e frequenti cambi d'intensità, e un suono etichettato come “upo” (pensato per risultare congruente alle immagini arrotondate), più omogeneo e lineare.

21 Per approfondire l'inibizione di ritorno si vedano le ricerche di Posner e Klein riportate nell'indice bibliografico, [\[68\]](#)[\[109\]](#)[\[110\]](#).

L'intero test si sviluppava in un totale di 462 prove, due terzi delle quali con la simultanea presenza di stimoli visivi e uditivi, mentre il rimanente terzo con soli stimoli visivi; per tutte le prove gli stimoli rilevanti ai quali era chiesto di porre attenzione erano quelli visivi. I soggetti erano collocati a 57 cm di distanza rispetto al monitor nel quale venivano mostrati gli stimoli visivi, dietro al quale venivano emessi gli stimoli uditivi attraverso uno speaker; l'intensità del suono era impostata a 66 dB.

Nei due terzi delle prove con entrambe le categorie di stimoli, ai partecipanti veniva mostrata per una durata di 80 ms una prima immagine (etichettata come *memorandum*), appartenente in modo randomico ad una famiglia di figure o all'altra (appuntite/smussate); tale stimolo visivo poteva essere preceduto (-200 ms), seguito (200 ms) oppure essere simultaneo (0 ms) ad uno stimolo uditivo (congruente o incongruente rispetto all'immagine). Dopo un intervallo variabile tra 300 e 1300 ms veniva mostrato un secondo stimolo visivo (etichettato come *probe*), congruente rispetto al primo (cioè appartenente alla stessa categoria). Nelle prove con i soli stimoli visivi invece, tra la prima e la seconda figura veniva mostrata l'immagine di disturbo per una durata di 100 ms. I partecipanti avevano infine un massimo di 2200 ms per stabilire se la seconda immagine in sequenza era uguale o meno alla precedente; era poi richiesto di indicare con un voto da 0 a 4 l'indice di sicurezza con il quale esprimevano la loro risposta.

I risultati ottenuti dalle prove con stimoli cross-modali audio-visivi sono stati messi a confronto con quelli relativi alle prove con i soli stimoli visivi.

In primo luogo, la presentazione anticipata dello stimolo uditivo (-200 ms) rispetto alla doppia presentazione degli stimoli visivi (suono[c/i]-*memorandum-probe*) non ha avuto alcun effetto sulla performance dei partecipanti; l'attenzione e la memoria visiva a breve termine esclude dunque l'interferenza di stimoli uditivi percepiti precedentemente rispetto alla successiva sequenza di stimoli visivi.

In secondo luogo, la presentazione simultanea (delay pari a 0 ms) di stimoli uditivi e visivi congruenti ha migliorato la performance dei partecipanti nelle prove con la seconda immagine presentata con un ritardo di 1300 ms (suono[c]/*memorandum-probe*[1300ms]), mentre risulta peggiore per entrambi i tempi di ritardo con stimoli incongruenti (suono[i]/*memorandum-probe*[300, 1300ms]); questo dimostrerebbe che la presenza simultanea di stimoli percepiti come congruenti possa risultare un rinforzo per l'integrazione cross-modale di stimoli appartenenti alle diverse modalità percettive, mentre stimoli incongruenti darebbero luogo a effetti di interferenza.

Inoltre, gli stimoli uditivi congruenti presentati con un ritardo di 200 ms rispetto al primo stimolo visivo hanno migliorato la performance dei partecipanti nelle prove con la seconda immagine ritardata di 300 ms (*memorandum-suono[c]-probe[300ms]*). Tale dato significa che la percezione cross-modale di stimoli uditivi congruenti percepiti durante una rapida successione di stimoli visivi vengono uniti assieme e associati ad un singolo evento (*crossmodal binding problem*).

Infine, stimoli uditivi incongruenti presentati 200 ms dopo la prima immagine hanno indebolito la performance delle prove in cui la seconda immagine veniva presentata con un ritardo di 1300 ms (*memorandum-suono[i]-probe[1300ms]*), interrompendo così la corretta elaborazione degli stimoli visivi e interferendo con l'attenzione visiva.

Conclusioni

Nel corso di questo scritto sono state accorpate e presentate in modo organico tutte le principali scoperte nel campo della percezione cross-modale, circoscrivendo l'attenzione sui rapporti tra i canali sensoriali uditivo e visivo.

Si è innanzitutto compresa l'importanza che quotidianamente ricopre la percezione cross-modale, e come essa sia costituita da un groviglio di funzioni sia innate sia acquisite da ogni individuo. È stata fatta luce sulle interazioni percettive tra la modalità uditiva e quella visiva, strutturando una prima tassonomia delle principali coppie di *features* appartenenti rispettivamente all'uno e all'altro canale sensoriale. La disanima degli studi più recenti mostra come l'interesse su tale tema sia tutt'ora molto alto; le ultime ricerche sperimentali stanno infatti portando avanti in modo scrupoloso la solida base posta negli anni sessanta e settanta del Novecento.

Per quanto concerne l'annoso dibattito tra le diverse posizioni in merito al rapporto tra sinestesia e percezione cross-modale, permangono due versioni contrastanti: chi vede la sinestesia come un fenomeno atipico e a sé stante, chi invece sostiene che, di base, sinesteti e non condividano le stesse funzioni percettive, solo in proporzioni diverse. Dunque, non volendo sposare nessuna delle due cause, le si prendano per buone entrambe, anche perché una versione non ne esclude per forza l'altra. Da una parte infatti è stato dimostrato che neurologicamente i soggetti sinesteti mostrano una struttura corticale in minima parte diversa rispetto ai soggetti non-sinesteti; inoltre l'attivazione e la consistenza degli effetti sinestesici mostra in modo chiaro dei tratti diversi rispetto ai comportamenti percettivi cross-modali più diffusi. Dall'altra parte però le due esperienze mostrano un'indubbia somiglianza e, per certi versi, possono essere paragonate e studiate assieme, ottenendo dei risultati per molti aspetti simili; inoltre, sapere che le esperienze dei soggetti sinesteti possono essere apprezzate e giudicate come altamente congruenti anche dai non-sinesteti [144] è indice di una similitudine di fondo.

Dove abbiano origine tali corrispondenze tra i diversi canali percettivi non è ancora stato stabilito con certezza; opinioni diverse stanno alimentando un dibattito ancora aperto, e in questa sede non è opportuno addentrarvisi. Si basti rimarcare che l'influenza linguistica potrebbe essere una delle concuse all'origine di tali

corrispondenze; l'utilizzo molto diffuso di metafore sinestetiche per definire i tratti degli stimoli di un certo canale sensoriale prendendo a prestito gli aggettivi per descrivere gli stimoli tipici di altri canali sensoriali è già di per sé un dato di non poco conto. Ma il fatto che le corrispondenze cross-modali siano state verificate già in età pre-linguistica è un ulteriore dato che non può essere sottovalutato, e suggerisce che i processi percettivi cross-modali siano, almeno in parte, insiti nella natura umana. Ci sono poi studi che hanno dimostrato come le metafore sinestetiche non sempre siano condivise da soggetti appartenenti a diversi gruppi linguistici [120]. È altamente probabile che tali processi cognitivi siano originati anche dall'esperienza che ognuno matura in rapporto al mondo fenomenico; per cui, ad esempio, i suoni acuti sono maggiormente congruenti alle posizioni più elevate nello spazio, così come i suoni gravi vengono associati a forme dalle grandi dimensioni. Secondo una prospettiva ecologica gibsoniana, le *affordances* che caratterizzano il mutuo rapporto tra animale e ambiente [53], si sedimenterebbero dunque nella memoria, assumendo così la funzione di “dizionario percettivo” utile a migliorare la ricostruzione di oggetti ed eventi unitari partendo da stimoli unidimensionali (ad es., pitch, loudness, timbro).

Alcune qualità sensoriali oggi risultano molto approfondite e costituiscono una prima base certa per poter parlare in modo compiuto di corrispondenze cross-modali, senza correre il rischio di incappare in associazioni empiriche di dubbia natura. Il sistema cognitivo umano è infatti in grado di percepire come congruenti o incongruenti certe coppie di stimoli audio-visivi, questo è un dato certo e ampiamente documentato. La simultanea percezione di segnali uditivi e visivi considerati congruenti tra loro tende a migliorare o amplificare i processi cognitivi, mentre stimolazioni incongruenti porterebbe ad una loro attenuazione. Pitch e loudness sono i tratti del suono più rappresentabili da un punto di vista visivo; anche il timbro lo è, sebbene ulteriori indagini sino auspicabili.

In linea generale, ci sono chiari indizi che ogni classe di corrispondenze cross-modali sia basato su un rapporto tra gli opposti: grande-piccolo, alto-basso, buio-luce, arrotondato-appuntito, ecc. Ciò rende scalabile la parametrizzazione delle qualità percettive a seconda dei casi: un suono verrà giudicato elevato da un punto di vista spaziale qualora, in una data circostanza, esso sia cognitivamente messo a confronto con una serie di altri suoni aventi ognuno una frequenza minore; ma lo stesso stimolo uditorio potrà essere meglio rappresentato da una bassa posizione spaziale se gli elementi acustici a disposizione sono altri e con altezze più acute. Le corrispondenze sono dunque determinate anche dal contesto, sebbene i tratti generali rimangano alterati.

I molti studi sperimentali hanno tuttavia lasciato un vuoto da coprire: la comprensione delle funzioni cross-modali in rapporto a stimoli audio-visivi variabili nelle loro qualità percettive. Il maggior numero dei test è infatti basato su stimoli con proprietà stazionarie. Quali sono, ad esempio, le *features* visive che meglio descrivono uno stimolo acustico la cui intensità, inizialmente nulla, cresce in modo costante fino ad un picco massimo e poi torna a zero? E quelle di un suono la cui altezza varia alternativamente da una frequenza acuta e una grave? Queste domande potrebbero trovare delle risposte empiriche basate sui modelli di base approfonditi in questo scritto; ma appositi studi potrebbero essere decisivi per appurare la veridicità o meno di tali risposte empiriche. È infatti già stato verificato nel rapporto tra pitch uditorio e dimensione visiva che stimoli variabili in modo continuo possono addirittura invertire il modello base della corrispondenza cross-modale studiato con stimoli stazionari.

Sarebbe inoltre interessante e potenzialmente utile indagare sulle possibili corrispondenze cross-modali scaturite da stimoli multipli, cioè ad esempio come siano, da un punto di vista percettivo, visivamente rappresentabili gruppi di suoni simultanei a differenti altezze (accordi o pattern sonori). È abbastanza intuitivo però che questi modelli strutturalmente complessi sono completamente diversi rispetto a quelli analizzati in questa sede e potrebbero quindi richiedere altri metodi di studio, derivati ad esempio dalla psicologia della forma (Gestalt).

C'è da tenere a mente che sia in natura sia in musica, difficilmente si trova traccia di singoli stimoli uditori stazionari; al contrario, solitamente vi è una complessità e variabilità che non ha nulla a che vedere con gli stimoli impiegati nei test sperimentali per lo studio della percezione cross-modale. La semplificazione è da ritenersi tuttavia necessaria per l'approfondimento e la completa comprensione dei meccanismi percettivi di più basso livello, ma in futuro, una volta chiariti in modo definitivo tali processi di basso livello, sarà necessario l'impiego di stimoli audio-visivi complessi e variabili. Anche allora potrebbe essere utile lo studio delle esperienze sinestetiche dei soggetti con forme di audizione colorata, poiché potrebbe rivelare la struttura delle mappature di alto livello tra la modalità uditoria e visiva condivise e apprezzate da tutti, così com'è stato verificato per le corrispondenze cross-modali di basso livello.

A prescindere dagli sviluppi futuri, ciò che è stato riportato in questo studio è comunque un utile punto di partenza per una serie di applicazioni sia tecniche sia artistiche. Lo studio della sinestesia e dell'integrazione multimodale offre il potenziale per ampliare queste esplorazioni verso un approccio olistico alla comprensione

scientifica dell'arte multimediale e del sound design. Molti sono gli esempi di applicazioni software che hanno unificato in un'interfaccia grafica la gestione e la modellazione del suono, dal *MetaSynth*²² di Éric Wenger al più noto UPIC (*Unité Polyagogique Informatique du CEMAMu*) di Iannis Xenakis [148], ma solo recentemente è stato studiato un modello di rappresentazione grafica fondato proprio sulle corrispondenze cross-modali; un primo esempio è rappresentato dal *Sound Mosaics* di Kostas Giannakis [51][52]. Altri progetti si sono cimentati nella rappresentazione visiva delle qualità percettive del suono, sperimentando la possibilità di organizzare e poter visualizzare una collezione di campionamenti attraverso interfacce grafiche intuitive, così come proposto da Grill e Flexer [56].

Infine, le conoscenze dei meccanismi percettivi cross-modali possono offrire utili spunti anche allo *Human-Computer Interaction* (HCI), cioè a quel campo di studi che indaga l'ottimizzazione dei meccanismi di interazione tra l'uomo e i prodotti digitali. Per quanto concerne il suono, l'interazione uomo-computer avviene attraverso le cosiddette *auditory interfaces* (interfacce uditive) [50], le quali sono essenzialmente di due tipi: *earcons*, suoni astratti che accompagnano i diversi eventi informatici (segnali di avviso), e le *auditoris icon*, ovvero imitazioni o rappresentazioni dei suoni di tutti i giorni (*everyday sounds*) associati alle interazioni tra uomo-macchina [39]. Per entrambe le categorie, un approfondimento sui processi cognitivi cross-modali può senz'altro facilitare la progettazione di *auditory interfaces* intuitive, che tengano conto cioè delle varie associazioni che implicitamente l'uomo adotta nel rapporto tra i suoni e le immagini.

22 <http://www.uisoftware.com/MetaSynth/>

Bibliografia

- [1] Adeli, Mohammad; Rouat, Jean; Molotchnikoff, Stéphane. *Audiovisual correspondence between musical timbre and visual shapes*, «Frontiers in Human Neuroscience», vol. 8, no. 352, 2014, pp. 1-11.
- [2] Balconi, Michela; Carrera, Alba. *Percezione cross-modale delle emozioni. Sincronizzazione di codici comunicativi*. In: *Neuropsicologia della comunicazione*, Springer, Milano, 2008.
- [3] Bankieris, Kaitlyn; Simner, Julia. *What is link between synesthesia and sound symbolism?*, «Cognition», vol. 136, March 2015, pp. 186-195.
- [4] Baron-Cohen, Simon et al. *Coloured speech perception: Is synesthesia what happens when modularity breaks down?*, «Perception», vol. 22, no. 4, 1993, pp. 419-426.
- [5] Bavelier, Daphne; Neville, Helen J. *Cross-Modal Plasticity: Where and How?*, «Nature Reviews Neuroscience», vol. 3, no. 6, 2002, pp. 443-452.
- [6] Beeli, Gian; Esslen, Michaela; Jäncke, Lutz. *Synesthesia: When coloured sounds taste sweet*, «Frontiers in Human Neuroscience», vol. 434, 2005, p. 38.
- [7] Belkin, Kira et al. *Auditory Pitch as a Perceptual Analogue to Odor Quality*, «Psychological Science», vol. 8, no. 4, 1997, pp. 340-342.
- [8] Ben-Artzi, Elisheva; Marks, Lawrence E. *Visual-auditory interaction in speeded classification: Role of stimulus difference*, «Perception & Psychophysics», vol. 57, no. 8, 1995, pp. 1151-1162.
- [9] Bernstein, Ira H.; Edelstein, Barry A. *Effects of some variations in auditory input upon visual choice reaction time*, «Journal of Experimental Psychology», vol. 87, no. 2, 1971, pp. 241-247.

- [10] Bernstein, Ira H.; Edelstein, Barry A. *Effects of some variations in auditory input upon visual choice reaction time*, «Journal of Experimental Psychology», vol. 87, no. 2, 1971, pp. 241-247.
- [11] Bertelson, Paul; Aschersleben, Gisa. *Automatic visual bias of perceived auditory location*, «Psychonomic Bulletin & Review», vol. 5, no. 3, 1998, pp. 482-489.
- [12] Bertelson, Paul et al. *The ventriloquist effect does not depend on the direction of deliberate visual attention*, «Perception & Psychophysics», vol. 62, no. 2, 2000, pp. 321-332.
- [13] Bismarck, G. von. Sharpness as an attribute of the timbre of steady sounds, «Acustica», vol. 30, no. 3, 1974, pp. 159-172.
- [14] Bond, Barbara; Stevens, Stanley S. Cross-modality matching of brightness to loudness by 5-years-old, «Perception & Psychophysics», vol. 6, no. 6, 1969, pp. 337-339.
- [15] Bonetti, Leonardo; Costa, Marco. *Pitch-verticality and pitch-size cross-modal interactions*, «Psychology of Music», 2017, pp. 1-17.
- [16] Bragança, G. F. F.; Fonseca, J. G. M.; Caramelli, P. *Synesthesia and music perception*, «Dement Neuropsychol», vol. 9, no. 1, 2015, pp. 16-23.
- [17] Bregman, Albert S.; Campbell, Jeffrey. *Primary auditory stream segregation and perception of order in rapid sequences of tones*, «Journal of Experimental Psychology», vol. 89, 1971, pp. 244-249.
- [18] Bregman, Albert S.; Achim, André. *Visual stream segregation*, «Perception & Psychophysics», vol. 13, no. 3, 1973, pp. 451-454.
- [19] Bregman, Albert S. *Auditory scene analysis: the perceptual organization of sound*, The MIT Press, Cambridge (Mass.), 1994.
- [20] Bremner, Andrew J. Et al. *“Bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches*

to Westerners, «Cognition», vol. 126, no. 2, 2013, pp. 165-172.

- [21] Bronner, Kai et al. *What Is the Sound of Citrus? Research on the Correspondences between the Perception of Sound and Flavour*, «Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM)», Thessaloniki, Greece, July 23-28 2012, pp. 142-148.
- [22] Bruzzi, Elena et al. *A SMARC Effect for Loudness*, «i-Perception», vol. 8, no. 6, Nov-Dec 2017, pp. 1-11.
- [23] Caivano, José Luis. *Color and Sound: Physical and Psychophysical Relations*, «Color Research & Application», vol. 19, no. 2, 1994, pp. 126-133.
- [24] Campen, Crétien van. *Artistic and Psychological Experiments with Synesthesia*, «Leonardo», vol. 32, no. 1, 1999, pp. 9-14.
- [25] Carello, Claudia; Anderson, Krista L.; Kundler-Peck, Andrew J. *Perception of Object Length by Sound*, «Psychological Science», vol. 9, no. 3, 1998, pp. 211-214.
- [26] Chang, Seah; Cho, Yang Seok. *Polarity correspondence effect between loudness and lateralized response set*, «Frontiers in Psychology», vol. 6, no. 683, 2015.
- [27] Chiou, Rocco; Stelter, Marleen; Rich, Anina N. *Beyond colour perception: Auditoryvisual synesthesia induces experiences of geometric objects in specific locations*, «Cortex», vol. 49, no. 6, 2013, pp. 1750-1763.
- [28] Cho, By et al. *Referential coding contributes to the horizontal SMARC effect*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 38, no. 3, 2012, pp. 726-734.
- [29] Coward, Sean W.; Stevens, Catherine J. *Extracting meaning from sound: Nomadic mappings, everyday listening, and perceiving object size from*

frequency, «The Psychological Record», vol. 54, 2004, pp. 349-364.

- [30] Davis, Edward R. *The fitness of names to drawings: A cross-cultural study in Tanganyika*, «British Journal of Psychology», vol. 52, 1961, pp. 259-268.
- [31] Day, Sean. *Trends in Synesthetically Colored Graphemes and Phonemes - - 2004 revision*, liberamente scaricabile al seguente url:
<http://www.daysyn.com/Day2004Trends.pdf>
- [32] Dolscheid, Sarah et al. *Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures*, «Psychological Science», vol. 25, no. 6, 2014, pp. 1256-1261.
- [33] Doovern, Anna et al. *Intrinsic Network Connectivity Reflects Consistency of Synesthetic Experiences*, «The Journal of Neuroscience», vol. 32, no. 22, 2012, pp. 7614-7621.
- [34] Eitan, Zohar et al. *Lower pitch is larger, yet falling pitches shrink: Interaction of pitch change and size change in speeded discrimination*, «Experimental Psychology», vol. 61, 2014, pp. 274-284.
- [35] Evans, Karla; Treisman, Anne. *Natural cross-modal mappings between visual and auditory features*, «Journal of Vision», vol. 10, no. 1, 2010, pp. 1-12.
- [36] Fastl, Hugo; Zwicker, Eberhard. *Psychoacoustics: Facts and Models*, Springer, Berlino, 3 ed., 2007.
- [37] Fernández-Prieto, Irune; Navarra, Jordi. *The higher the pitch the larger its crossmodal influence on visuospatial processing*, «Psychology of Music», vol. 45, no. 5, Nov-Dec 2017, pp. 713-724.
- [38] Fernandez-Prieto, Irune et al. *Does Language Influence the Vertical Representation of Auditory Pitch and Loudness?*, «i-Perception», vol. 8, no. 3, 2017.
- [39] Fernström, Mikael. *Sound objects and human-computer interaction*, in Rocchesso; Fontana (Edit.). *The Sounding Object*, Edizioni di Mondo Estremo,

Firenze, 2003, pp. 33-46.

- [40] Ferrero, Franco E.; Magno Caldognetto, Emanuela. *Elementi di fonetica acustica*, in Croatto (A cura di). *Trattato di foniatria e logopedia*, La Garangola, Padova, 1986, pp. 155-196.
- [41] Fletcher, Harvey; Munson, Wilden. *Loudness, Its Definition, Measurement and Calculation*, «The Journal of the Acoustical Society of America», vol. 5, no. 2, 1933, pp. 82-108.
- [42] Fornazzari, Luis et al. “*Blue is music to my ears*”: *Multimodal synesthesia after a thalamic stroke*, «*Neurocase*», vol. 18, no. 4, 2012, pp. 318-322.
- [43] Frassinetti, Francesca; Bolognini, Nadia; Làdavas, Elisabetta. *Enhancement of visual perception by crossmodal visuo-auditory interaction*, «*Experimental Brain Research*», vol. 147, no. 3, 2002, pp. 332-343.
- [44] Gal, David; Wheeler, S. Christian; Shiv, Baba. *Cross-Modal Influences on Gustatory Perception*. November 14, 2007.
- [45] Galeev, Bulat Maxmudovich. *The Nature and Functions of Synesthesia in Music*, «*Leonardo*», vol. 40, no. 3, 2007, pp. 285-288.
- [46] Gallace, Alberto; Spence, Charles. *Multisensory synesthetic interactions in the speeded classification of visual size*, «*Perception & Psychophysics*», vol. 68, no. 7, 2006, pp. 1191-1203.
- [47] Gandour, Jackson T. *The perception of tone*, in Fromkin (Edit.). *Tone: A Linguistic Survey*, Academic Press, New York, 1978, pp. 41-76.
- [48] Garner, Wendell R. *The processing of information and structure*, Lawrence Erlbaum Associates, New York, 1974.
- [49] Garner, Wendell. The Relationship between Colour and Music, «*Leonardo*», vol. 11, no. 3, 1978, pp. 225-226.
- [50] Gaver, William W. *Auditory Interfaces*, in Helander; Landauer; Prabhu (Edit.).

Handbook of Human-Computer Interaction, Elsevier Science, Amsterdam, 1997, pp. 1003-1042.

- [51] Giannakis, Kostas. *A comparative evaluation of auditory-visual mappings for sound visualisation*, «Organised Sound», vol. 11, no. 3, 2006, pp. 297-307.
- [52] Giannakis, Kostas; Smith, Matt. *Imaging Soundscapes: Identifying Cognitive Associations between Auditory and Visual Dimensions*, in Rolf Inge Godoy, Harald Jorgensen (Edit.). *Musical imagery*, Swets & Zeitlinger, Lisse, 2001, pp. 161-179.
- [53] Gibson, James J. *Un approccio ecologico alla percezione visiva*, tr. di Riccardo Luccio, (tit. or. *The ecological approach to visual perception*, Houghton Mifflin company, Boston, 1979) Fabbri, Milano, 2007.
- [54] Ginsberg, Leon. *A case of Synesthesia*, «The American Journal of Psychology», vol. 34, no. 4, 1923, pp. 582-589.
- [55] Giordano, Bruno L.; Rocchesso, Davide; McAdams, Stephen. *Integration of Acoustical Information in the Perception of Impacted Sound Sources: The Role of Information Accuracy and Exploitability*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 36, no. 2, 2010, pp. 462-476.
- [56] Grill, Thomas; Flexer, Arthur. *Visualization of Perceptual Qualities in Textural Sounds*, «Proceedings of the International Computer Music Conference (ICMC 2012)», Ljubljana, Slovenia, September 2012, pp. 589-596.
- [57] Hamilton-Fletcher, Giles et al. *Sound Properties Associated With Equiluminant Colours*, «Multisensory research», vol. 30, no. 3-5, 2017, pp. 337-362.
- [58] Handel, Stephen. *Perceptual coherence : hearing and seeing*, Oxford University Press, Oxford, 2006.
- [59] Harrison, John; Baron-Cohen, Simon. *Synesthesia: An Account of Coloured*

- Hearing*, «Leonardo», vol. 27, no. 4, 1994, pp. 343-346.
- [60] Hartmann, Matthias; Mast, Fred W. *Loudness Counts: Interactions Between Loudness, Number Magnitude and Space*, «The Quarterly Journal of Experimental Psychology», vol. 70, no. 7, 2017, pp. 1305-1322.
- [61] Head, Phineas de Thornley. *Synesthesia: Pitch-Colour Isomorphism in RGB-Space?*, «Cortex», vol. 42, no. 2, 2006, pp. 164-174.
- [62] Hubbard, Timothy L. *Synesthesia-like Mappings of Lightness, Pitch, and Melodic Interval*, «The American Journal of Psychology», vol. 109, no. 2, 1996, pp. 219-238.
- [63] Imai, Mutsumi; Kita, Sotaro. *The sound symbolism bootstrapping hypothesis for language acquisition and language evolution*, «Philosophical Transactions of the Royal Society B», vol. 369, 2014.
- [64] Jamal, Yaseen et al. *Interactions Between Auditory Elevation, Auditory Pitch and Visual Elevation During Multisensory Perception*, «Multisensory research», 2017, Supplementary Material. Brill Online.
- [65] Jonas, Clare; Spiller, Mary Jane; Hibbard, Paul. *Summation of visual attributes in auditory-visual crossmodal correspondences*, «Psychonomic Bulletin & Review», vol. 24, no. 4, 2017, pp. 1104-1112.
- [66] Kandinsky, Wassily. *Lo spirituale nell'arte*, a cura di Elena Pontiggia, (tit. or. *Über das Geistige in der Kunst*), SE, Milano, 2005.
- [67] Kemp, Sarah; Gilbert, Avery. *Odor Intensity and Color Lightness Are Correlated Sensory Dimensions*, «The American Journal of Psychology», vol. 110, no. 1, 1997, pp. 35-46.
- [68] Klein, Raymond M. *Inhibition of return*, «Trends in Cognitive Sciences», vol. 4, no. 4, 2000, pp. 138-147.
- [69] Köhler, Wolfgang. *Akustische Untersuchungen II*, «Zeitschrift für

- Psychologie», vol. 58, 1910, pp. 59-140.
- [70] Köhler, Wolfgang. *La psicologia della Gestalt*, tr. di G. De Toni, (ed. or. *Gestalt Psychology*, Liveright Publishing Corporation, New York, 1947), Feltrinelli, Milano, 1981.
- [71] Kwok, Sze Chai et al. *A biphasic effect of cross-modal priming on visual shape recognition*, «Acta Psychologica», vol. 183, 2018, pp. 43-50.
- [72] Lacey, Simon et. al. *Synesthesia strengthens sound-symbolic cross-modal correspondences*, «European Journal of Neuroscience», vol. 44, no. 9, Nov. 2016, pp. 2716-2721.
- [73] Lewkowicz, David J.; Turkewitz, Gerald. *Cross-Modal Equivalence in Early Infancy: Auditory-Visual Intensity Matching*, «Developmental Psychology», no. 6, 1980, pp. 597-607.
- [74] Lidji, Pascale et al. *Spatial Associations for Musical Stimuli: A Piano in the Head?*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 33, no. 5, 2007, pp. 1189-1207.
- [75] Lipscomb, Scott D.; Kim, Eugene M. *Perceived match between visual parameters and auditory correlates: An experimental multimedia investigation*, «Proceedings of the 8th International Conference on Music Perception and Cognition (ICMPC)», Evanston, Illinois (USA), August 3-8 2004, pp. 142-148.
- [76] Marks, E. Lawrence; Stevens, Joseph C. *Individual brightness functions*, «Perception & Psychophysics», vol. 1, no. 1, 1966, pp. 17-24.
- [77] Marks, Lawrence E. *On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness*, «American Journal of Psychology», vol. 87, 1947, no. 1-2, pp. 173-188.
- [78] Marks, Lawrence E. *On Colored-Hearing Synesthesia: Cross-Modal Translations of Sensory Dimensions*, «Psychological Bulletin», vol. 82, no. 3,

June 1975, pp. 303-331.

- [79] Marks, Lawrence E. *The Unity of the Senses: Interrelations among the Modalities*, Academic Press, New York, San Francisco, London, 1978.
- [80] Marks, Lawrence E. *Bright Sneezes and Dark Coughs: Loud Sunlight and Soft Moonlight*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 8, no. 2, April 1982, pp. 177-193.
- [81] Marks, Lawrence E. *On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 13, no. 3, 1987, pp. 384-394.
- [82] Marks, Lawrence E. *On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 15, no. 3, 1989, pp. 586-602.
- [83] Marks, Lawrence E. *On Perceptual Metaphors*, «Metaphor and Symbolic Activity», vol. 11, no. 1, 1996, pp. 39-66.
- [84] Marks, Lawrence E. *Cross-Modal Interactions in Speeded Classification*, in Calvert; Spence; Stein (Edit.). *The Handbook of Multisensory Processes*, The MIT Press, Cambridge (Mass.), 2004, pp. 85-105.
- [85] Marks, Lawrence E.; Ben-Artzi, Elisheva; Lakatos, Stephen. *Cross-modal interactions in auditory and visual discrimination*, «International Journal of Psychophysiology», vol. 50, 2003, pp. 125-145.
- [86] Marks, Lawrence E.; Hammeal, Robin J.; Bornstein, Marc H. *Perceiving Similarity and Comprehending Metaphor*, «Monographs of the Society for Research in Child Development», vol. 52, no. 1, 1987, pp. 1-102.
- [87] Martinelli, Riccardo. *Teoria dei Suoni e Antropologia*, in Desideri; Matteucci (A cura di). *Estetiche della percezione*, Firenze University Press, Firenze, 2007, pp. 87-100.
- [88] Martino, Gail; Marks, Lawrence E. *Synesthesia: Strong and Weak*, «Current

Directions in Psychological Science», vol. 10, no. 2, April 2001, pp. 61-65.

- [89] Martino, Gail; Marks, Lawrence E. *Cross-modal interaction between vision and touch: the role of synesthetic correspondence*, «Perception», vol. 29, no. 6, 2000, pp. 745-754.
- [90] Maurer, Daphne; Pathman, Thanujeni; Mondloch, Catherine J. *The shape of boubas: sound-shape correspondences in toddlers and adults*, «Developmental Science», vol. 9, no. 3, 2006, pp. 316-322.
- [91] McGurk, Harry; MacDonald, John. *Hearing lips and seeing voices*, «Nature», vol. 264, 1976, pp. 746-748.
- [92] Melara, R. D.; O'Brien, T. P. *Interaction between synesthetically corresponding dimensions*, «Journal of Experimental Psychology: General», vol. 116, no. 4, 1987, pp. 323-336.
- [93] Mills, Carol B.; Boteler, Edith H.; Oliver, Glenda K. *Digit Synesthesia: A Case Study Using a Stroop-Type Test*, «Cognitive Neuropsychology», vol. 16, no. 2, 1980, pp. 181-191.
- [94] Mondloch, Catherine; Maurer, Daphne. *Do small white balls squeak? Pitch-object correspondences in young children*, «Cognitive, Affective, & Behavioral Neuroscience», vol. 4, no. 2, 2004, pp. 133-136.
- [95] Neufeld, J. et al. *The neural correlates of coloured music: A functional MRI investigation of auditory-visual synesthesia*, «Neuropsychologia», vol. 50, 2012, pp. 85-89.
- [96] Neuhoff, John G. *Interacting perceptual dimension*, «Ecological Psychoacoustics», 2004, pp. 249-269.
- [97] Nielsen, Alan; Rendall, Drew. *The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon*, «Canadian Journal of Experimental Psychology», vol. 65, no. 2, 2011, pp. 115-124.

- [98] Nielsen, Alan; Rendall, Drew. *Parsing the Role of Consonants Versus Vowels in the Classic Takete-Maluma Phenomenon*, «Canadian Journal of Experimental Psychology», vol. 67, no. 2, 2013, pp. 153-163.
- [99] O'Leary, Ann; Rhodes, Gillian. *Cross-modal effects on visual and auditory object perception*, «Perception & Psychophysics», vol. 35, no. 6, 1984, pp. 565-569.
- [100] O'Boyle, Michael W.; Tarte, Robert D. *Implications for Phonetic Symbolism: The Relationship Between Pure Tones and Geometric Figures*, «Journal of Psycholinguistic Research», vol. 9, no. 6, 1980, pp. 535-544.
- [101] Ozturk, Ozge; Krehm, Madelaine; Vouloumanos, Athena. *Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds*, «Journal of Experimental Child Psychology», vol. 114, no. 2, 2013, pp. 173-186.
- [102] Palladino, Pietro. *Manuale di illuminazione*, Tecniche nuove, Milano, 2005.
- [103] Parise, Cesare; Spence, Charles. “*When Birds of a Feather Flock Together*”: *Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes*, «PLoS ONE», vol. 4, no. 5, May 2009, <https://doi.org/10.1371/journal.pone.0005664>.
- [104] Parise, Cesare; Spence, Charles. *Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test*, «Experimental Brain Research», vol. 220, no. 3-4, 2012, pp. 319-333.
- [105] Parise, Cesare V.; Knorre, Katharina; Ernst, Marc O. *Natural auditory scene statistics shapes human spatial hearing*, «Proceedings of the National Academy of Sciences of the United States of America (PNAS)», vol. 111, no. 16, April 22 2014, pp. 6104-6108.
- [106] Pitteri, Marco et al. *Naturally together: pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians*, «Psychological Research», vol. 81, no. 1, 2017, pp. 243-254.

- [107] Pomerantz, J. R.; Pristach, E. A.; Carson, C. E. *Attention and object perception*, in Shepp; Ballesteros (Edit.). *Object perception: Structure and process*, Lawrence Erlbaum Associates, Hillsdale, New Jersey; Hove and London, 1989, pp. 53-89.
- [108] Posner, Michael. *Orienting of Attention*, «Quarterly Journal of Experimental Psychology», vol. 32, no. 1, 1980, pp. 3-25.
- [109] Posner, Michael; Cohen, Yoav. *Components of visual orienting*, «Attention and Performance», vol. 32, 1984, pp. 531-556.
- [110] Posner, Michael et al. *Inhibition of return: Neural Basis and Function*, «Cognitive Neuropsychology», vol. 2, no. 3, 1985, pp. 211-228.
- [111] Ramachandran, Vilayanur S.; Hubbard, Edward M. *Synaesthesia – a window into perception, thought and language*, «Journal of Consciousness Studies», vol. 8, no. 12, 2001, pp. 3-34.
- [112] Rusconi, Elena et al. *Spatial representation of pitch height: the SMARC effect*, «Cognition», vol. 99, no. 2, 2006, pp. 113-129.
- [113] Sacks, Oliver. *Musicofilia: racconti sulla musica e il cervello*, Adelphi, Milano, 2 ed., 2008.
- [114] Sakamoto, Maki; Watanabe, Junji. *Cross-Modal Associations between Sounds and Drink Tastes/Textures: A Study with Spontaneous Production of Sound-Symbolic Words*, «Chemical Senses», vol. 41, no. 3, 2016, pp. 197-203.
- [115] Schafer, R. Murray. *Il paesaggio sonoro*, tr. di Nemesio Ala, (tit. or. *The Tuning of the World*, Alfred A. Knopf, New York, 1979), Ricordi:LIM, Milano, 1985.
- [116] Schneider, Bruce A.; Bissett, Randall J. *The dimensions of tonal experience: A nonmetric multidimensional scaling approach*, «Perception & Psychophysics», vol. 30, no. 1, 1981, pp. 39-48.
- [117] Seo, Han-Seok et al. *Congruent Sound Can Modulate Odor Pleasantness*,

«Chemical Senses», vol. 39, no. 3, 2014, pp. 215-228.

- [118] Shams, Ladan; Kamitani, Yukiyasu; Shimojo, Shinsuke. *Visual illusion induced by sound*, «Cognitive Brain Research», vol. 14, 2002, pp. 147-152.
- [119] Shams, Ladan; Kamitani, Yukiyasu; Shimojo, Shinsuke. *Illusion: What you see is what you hear*, «Nature», vol. 408, 2000, p. 788.
- [120] Shang, Nan; Styles, Suzy J. *Is a High Tone Pointy? Speakers of Different Languages Match Mandarin Chinese Tones to Visual Shapes Differently*, «Frontiers in Psychology», vol. 8, no. 2139, 2017.
- [121] Shipley, Thorne. *Auditory Flutter-Driving of Visual Flicker*, «Science», vol. 145, no. 3638, 1964, pp. 1328-1330.
- [122] Simner, Julia; Cuskley, Christine; Kirby, Simon. *What sound does taste? Cross-modal mappings across gustation and audition*, «Perception», vol. 39, 2010, pp. 553-569.
- [123] Singh, Leher; Fu, Charlene S. L. *A New View of Language Development: The Acquisition of Lexical Tone*, «Child Development», vol. 87, no. 3, 2016, pp. 834-854.
- [124] Smalley, Denis. *Space-form and the acousmatic image*, «Organised Sound», vol. 12, no. 1, 2007, pp. 35-58.
- [125] Smith, Linda B.; Sera, Maria D. *A Developmental Analysis of the Polar Structure of Dimensions*, «Cognitive Psychology», vol. 24, 1992, pp. 99-142.
- [126] Spence, Charles. *Audiovisual multisensory integration*, «Acoustical Science & Technology», vol. 28, no. 2, 2007, pp. 61-70.
- [127] Spence, Charles. *Crossmodal correspondence: A tutorial review*, «Attention, Perception, & Psychophysics», vol. 73, no. 4, 2011, pp. 971-995.
- [128] Stein, Barry E. et al. *Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis*, «Journal of Cognitive Neuroscience», vol.

8, no. 6, 1996, pp. 497-506.

- [129] Stein, Barry E. *Neural mechanism for synthesizing sensory information and producing adaptive behaviors*, «Experimental Brain Research», vol. 123, no. 1-2, October 1998, pp. 124-135.
- [130] Stevens, S. S. *The attributes of tones*, «Proceedings of the National Academy of Sciences», vol. 20, 1934, pp. 457-459.
- [131] Stevens, Stanley S. *The Volume and Intensity of Tones*, «The American Journal of Psychology», vol. 46, no. 3, 1934, pp. 397-408.
- [132] Stumpf, Carl. *Tonpsychologie I*, Hirzel, Leipzig, 1883.
- [133] Takeshima, Yasuhiro; Gyoba, Jiro. *High-intensity sound increases the size of visually perceived objects*, «Attention, Perception, & Psychophysics», vol. 75, no. 3, 2013, pp. 501-507.
- [134] Tarte, Robert D.; Barritt, Loren S. *Phonetic Symbolism in Adult Native Speakers of English: Three Studies*, «Language and Speech», vol. 14, no. 2, 1971, pp. 158-168.
- [135] Tarte, Robert D. *Phonetic Symbolism in Adult Native Speakers of Czech*, «Language and Speech», vol. 17, no. 1, 1974, pp. 87-94.
- [136] Tonelli, Alessia; Cuturi, Luigi F.; Gori, Monica. *The Influence of Auditory Information on Visual Size Adaptation*, «Frontiers in Neuroscience», vol. 11, no. 594, 2017.
- [137] Tulving, Endel; Lindsay, Paul. *Identification of simultaneously presented simple visual and auditory stimuli*, «Acta Psychologica», vol. 27, 1967, pp. 101-109.
- [138] Tzanetakis, George; Cook, Perry. *Multifeature Audio Segmentation for Browsing and Annotation*, «Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics », New York, 20-20

Oct. 1999.

- [139] Vatakis, Argiro; Ghazanfar, Asif A.; Spence, Charles. *Facilitation of multisensory integration by the “unity effect” reveals that speech is special*, «Journal of Vision», vol. 8, no. 9, 2008, pp. 1-11.
- [140] Vroomen, Jean; De Gelder, Beatrice. *Sound Enhances Visual Perception: Cross-modal Effects of Auditory Organization on Vision*, «Journal of Experimental Psychology: Human Perception and Performance», vol. 26, no. 5, 2000, pp. 1583-1590.
- [141] Walker, Peter et al. *Preverbal Infants’ Sensitivity to Synaesthetic Cross-Modality Correspondences*, «Psychological Science», vol. 21, no. 1, 2010, pp. 21-25.
- [142] Walker, Robert. *The effects of culture, environment, age, and musical training on choices of visual metaphors for sound*, «Perception & Psychophysics», vol. 42, no. 5, 1987, pp. 491-502.
- [143] Ward, Jamie; Huckstep, Brett; Tsakanikos, Elias. *Sound-Colour Synaesthesia: to what extent does it use cross-modal mechanisms common to us all?*, «Cortex», vol. 42, no. 2, 2006, pp. 264-280.
- [144] Ward, Jamie et al. *The aesthetic appeal of auditory-visual synaesthetic perceptions in people without synaesthesia*, «Perception», vol. 37, no. 8, 2008, pp. 1285-1296.
- [145] Welch, Robert B.; DuttonHurt, Lance D.; Warren, David H. *Contributions of audition and vision to temporal rate perception*, «Perception & Psychophysics», vol. 39, no. 4, 1986, pp. 294-300.
- [146] Wells, Alan. *Music and Visual Color: A Proposed Correlation*, «Leonardo», vol. 13, no. 2, 1980, pp. 101-107.
- [147] Westbury, Chris. *Implicit sound symbolism in lexical access: Evidence from an interference task*, «Brain and Language», vol. 93, no. 1, 2005, pp. 10-19.

- [148] Xenakis, Iannis. *Formalized Music*, Pendragon Press, New York, 1992.
- [149] Yau, Jeffrey M. et al. *Temporal Frequency Channels are Linked Across Audition and Touch*, «Current Biology», vol. 19, 2009, pp. 561-566.
- [150] Zamm, Anna et al. *Pathways to seeing music: Enhanced structural connectivity in colored-music synesthesia*, «NeuroImage», vol. 74, no. 1, 2013, pp. 359-366.
- [151] Zeng, Hang et al. *The source of visual size adaptation*, «Journal of Vision», vol. 17, no. 8, 2017, pp. 1-15.

Elenco delle figure

| | |
|--|----|
| Illustration 1: Curve isofoniche di Fletcher e Munson..... | 24 |
| Illustration 2: <i>Sistema Munsell dei colori</i> | 25 |
| Illustration 3: Modello base per la corrispondenza cross-modale pitch-luminosità..... | 26 |
| Illustration 4: Modello base per la corrispondenza cross-modale pitch-dimensione..... | 28 |
| Illustration 5: Modello base per la corrispondenza cross-modale pitch-elevazione..... | 30 |
| Illustration 6: Figure geometriche utilizzate nel test di O'Boyle e Tarte..... | 35 |
| Illustration 7: Modello base per la corrispondenza cross-modale pitch-geometria..... | 35 |
| Illustration 8: Modello base per la corrispondenza cross-modale loudness-luminosità..... | 36 |
| Illustration 9: Modello base per la corrispondenza cross-modale loudness-dimensione..... | 38 |
| Illustration 10: Maluma e Takete..... | 40 |
| Illustration 11: Bouba e Kiki..... | 40 |
| Illustration 12: Corrispondenza uditivo-visiva tra suoni puri e diversi profili..... | 41 |
| Illustration 13: Modello base per la corrispondenza cross-modale fra timbro sonoro e profilo della forma..... | 42 |
| Illustration 14: Segmentazione visiva con quattro punti luminosi, illuminati alternativamente in quest'ordine: 1, 2, 3, 4 a bassa e alta velocità..... | 44 |
| Illustration 15: Differenti segmentazioni dello stream audio a bassa e alta frequenza..... | 44 |
| Illustration 16: Rapporto tra la frequenza sonora e la "frequenza del colore" proposto da Garner (1978)..... | 47 |
| Illustration 17: Rapporto tra luminosità del colore e loudness sonora; José Caivano..... | 48 |
| Illustration 18: Rapporto tra tonalità (chroma) del colore e timbro sonoro; José Caivano..... | 49 |
| Illustration 19: Inflessioni dei quattro toni cinesi..... | 67 |
| Illustration 20: Stimoli visivi impiegati nel test..... | 68 |