

# TRATTI FONOLOGICI, OGGETTI UDITIVI E ILLUSIONI<sup>1</sup>

Relazione a invito di Sarah Hawkins

Traduzione e riduzione a cura di  
Annachiara Varriale e Renata Savy

## 1. INTRODUZIONE

La definizione dei tratti distintivi è una definizione problematica: essi si configurano come ‘entità fonologiche astratte’ (Jakobson et alii, 1952; Chomsky & Halle, 1968) con specifiche proprietà fonetiche, difficili tuttavia da identificare nel segnale (cfr. Kingston & Diehl, 1994; Stevens, 2002), soprattutto nel parlato spontaneo. Tra il tratto fonologico e i suoi correlati fisici non esistono, infatti, rapporti permanenti di tipo biunivoco: in genere una categoria fonologica è segnalata da una co-occorrenza complessa di proprietà acustiche, le cui caratteristiche spettro-temporali sono tutt’altro che semplici da evidenziare. D’altro canto, l’occorrenza nel segnale di correlati fonetici di un determinato tratto fonologico non necessariamente è indice della presenza di quel tratto in una unità linguistica di qualsivoglia livello di analisi.

Con buona probabilità, le ragioni di tale difficoltà descrittiva e analitica non sono da attribuire né all’eventuale inadeguatezza dell’inventario dei tratti distintivi, né alla loro relazione con il segnale, ma sono da ricercarsi nel fatto che i processi percettivi uditivi sono più attivi e costruttivi di quanto si pensi.

Studi recenti sulla percezione visiva (Roy, 2005a, 2005b; Sprague et alii, 2007) hanno dimostrato che essa si configura come processo attivo in cui lo stimolo sensoriale induce un’interpretazione che si basa contemporaneamente sull’esperienza e sulla memoria di eventi passati. Nonostante sia stato largamente ipotizzato (Halle & Stevens, 1962; Stevens & Halle, 1967; Fowler, 1986) che la percezione uditiva funzioni in maniera molto simile, il dibattito si è concentrato sulla scelta tra modelli motori vs uditivi, mentre è stato trascurato il ruolo centrale dell’interazione tra memoria e sensazione.

Questo lavoro parte invece dall’assunto che i processi cerebrali coinvolti nella percezione sono fondamentalmente indipendenti dalla modalità sensoriale (J. Hawkins 2004): alcuni paralleli tra illusioni uditive e illusioni ottiche dimostrano che esiste una mediazione tra “realtà” percepita, aspettative e caratteristiche fisiche del segnale.

## 2. CORRELATI ACUSTICI E PERCETTIVI DI TRATTI DISTINTIVI

In questo paragrafo vengono discussi, a titolo di esempio, i correlati acustici e percettivi di due tratti fonologici distintivi, al fine di mostrare la difficoltà di stabilire un rapporto biunivoco tra unità linguistiche e caratteristiche fisiche, punto di partenza per le riflessioni di questo lavoro.

---

<sup>1</sup> Il testo qui presentato è un sunto del lavoro “Phonological features, auditory objects, and illusions”, pubblicato in *Journal of Phonetics* 38, 1, January 2010, Pages 60-89, al quale si rimanda per i dettagli del contenuto, le immagini e i riferimenti bibliografici completi. Si ringrazia Marco Aldo Piccolino-Boniforti per la revisione del testo italiano.

### 2.1. Il tratto [ $\pm$ sonoro]

Il correlato acustico principale del tratto [ $\pm$ sonoro] è, in linea teorica, semplice da identificare e consiste nell'assenza o presenza di barra sonora. Tuttavia, molte differenti proprietà contribuiscono alla produzione e alla percezione del tratto in sé.

La Figura 1 mostra una serie di tratti fonetici associati a consonanti occlusive intervocaliche sorde e sonore:

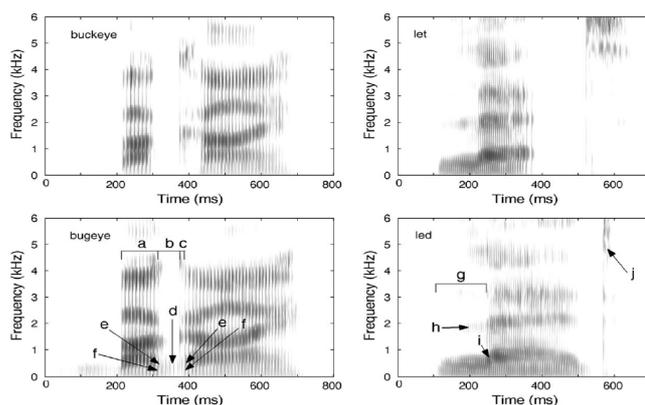


Figura 1: a sinistra, sonagrammi di *buckeye* e *bugeye*; a destra, sonagrammi di *let* e *led*.

Come si nota dai sonagrammi, la covariazione di queste proprietà produce **sillabe** caratterizzate da significativa presenza di energia periodica a bassa frequenza in contesto con occlusive sonore, e significativa presenza di energia aperiodica, silenzio ed attività in alta frequenza in contesto con occlusive sorde. Inoltre, la presenza di sonorità implica:

- a) nucleo sillabico precedente l'occlusione più lungo;
- b) occlusione più corta seguita da un
- c) VOT più breve;
- d) barra sonora più intensa durante l'occlusione;
- e) F1 più bassa ai confini dell'occlusione;
- f)  $f_0$  più bassa ai confini dell'occlusione
- g) consonante in testa di sillaba più lunga;
- h) transizioni formantiche più pronunciate al rilascio della consonante in testa;
- i) ampiezza dell'energia aperiodica più bassa al rilascio dell'ostruente in coda.

Si può dimostrare che le proprietà fisiche si combinano per produrre un 'percepto robusto' (Hawkins & Nguyen 2004) del tratto [ $\pm$ sonoro], ma è noto che a) non è necessaria la presenza di tutte; b) esiste un effetto di compensazione tra valori più pronunciati di un tratto acustico e valori più ambigui di un altro; c) i diversi tratti acustici hanno differente rilevanza percettiva; e d) tale gerarchia percettiva è diversa da lingua a lingua.

### 2.2. Il tratto [ $\pm$ nasale]

Le consonanti nasali in onsets di sillaba o in posizione intervocalica mostrano i seguenti tratti tipici:

- a) brusca riduzione dell'ampiezza generale del segnale all'abbassamento del velo, e repentino aumento al sollevamento;
- b) generale stabilità dell'ampiezza della forma d'onda;
- c) forma d'onda relativamente semplice caratterizzata da formanti fortemente 'smorzate' (*damped*) le cui frequenze variano poco e sono accompagnate da antirisonanze (*spectral zeros*).

Le uniche variazioni di frequenza (rare nel parlato accurato) riguardano F2 nasale in coarticolazione con i segmenti adiacenti durante la fase di chiusura orale. Il murmure nasale stabile di bassa energia risulta un buon indice di nasalità, ma non è un buon parametro per l'identificazione del luogo di articolazione (non distingue quindi /m/ da /n/ e così via...).

Le occlusive nasali in coda sillabica (e tutte le nasali in parlato connesso, spontaneo o veloce) mostrano un'alta variabilità coarticolatoria che spesso riflette con una certa sistematicità attributi non fonologici del segnale, come, ad esempio, la funzione grammaticale o lessicale dell'item (Local, 2003): nella sequenza *I'm going to check when the thyme goes in* il segmento *I'm* (che svolge un ruolo grammaticale) può essere interessato da diversi gradi di riduzione nelle forme più informali e veloci di parlato, fino ad arrivare al completo trasferimento della nasalità alla vocale ([ɜ̃ ɡɔ̃nə 'tʃɛk]) o alla manifestazione di segmento nasale sillabico o traccia nasale con assimilazione del luogo di articolazione della consonante successiva ([ɪŋɡɔ̃nə 'tʃɛk] o [ʰɪŋɡɔ̃nə 'tʃɛk]). Nessuno dei due pattern si ritrova nella parola *thyme* (che ha funzione lessicale), a meno che il contesto non sia tanto ridotto da implicare una completa cancellazione di *I'm*.

Si realizza pertanto per il tratto [ $\pm$ nasale] similmente a quanto evidenziato per il tratto [ $\pm$ sonoro], una mancanza di biunivocità tra proprietà acustiche e tratto fonologico.

### 3. PERCEZIONE “CREATIVA” E CATEGORIZZAZIONE

Le unità dell'analisi linguistica hanno senso, dunque, come categorie potenziali di produzione e percezione, ma nessuna ha una definizione in termini acustici interamente esaustiva. E' pertanto legittimo continuare ad assumere tali unità discrete, a livelli linguistici differenti, come base dei processi del linguaggio, quando non riusciamo esaustivamente ad identificare o definire le unità stesse? Una tale posizione sembra giustificabile solo se si stabilisce che i processi di percezione uditiva si configurano come processi di “creazione” attiva di unità discrete a partire da proprietà dello stimolo sensoriale variabili ed estese su domini di tempo multipli; la loro classificazione è dunque totalmente dipendente dall'elaborazione attiva e determinata dal contesto.

I progressi nella ricerca sulla percezione visiva e sul ‘*brain imaging*’ hanno messo in luce che diverse elaborazioni neurologiche contribuiscono a produrre il senso della “tangibilità” di un oggetto in assenza di uno stimolo di tipo tattile. E' legittimo pensare che la percezione uditiva segua principi generali simili. Nei prossimi paragrafi vengono tracciati paralleli tra illusioni ottiche, illusioni acustiche con stimoli non linguistici e possibili estensioni alla percezione del parlato (in particolare dei tratti distintivi).

### 4. ILLUSIONI DI CONTRASTO

L'illusione dei Triangoli di Kanizsa (Kanizsa & Gerbino, 1982) mostra che la percezione di confini categoriali è indotta da effetti di *contrasto visivo*: tra le immagini in fig. 2, quelle con i confini più netti (a sinistra) inducono l'immediata percezione di un

triangolo; nella terza figura, la percezione interviene non appena l'occhio si abitua ad una risoluzione più bassa; nella figura più a destra, in cui i cambi sono più sfumati e gradualmente, è possibile ricostruire l'immagine anche a partire dai pochi punti (gli angoli esterni) in cui il contrasto è più brusco.

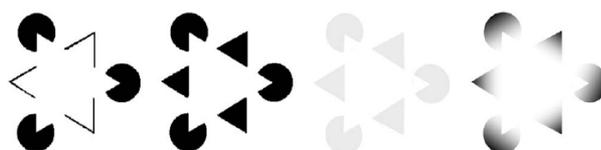


Figura 2: Triangoli di Kanizsa

Questa illusione può avere diversi paralleli con la percezione uditiva: ciò che conta non è la presenza di determinate proprietà fisiche, quanto i valori relativi tra i parametri di uno stimolo sensoriale che permettono al cervello di *costruire* il percepito più probabile.

L'esempio in fig. 3 è un *pattern* acustico assai comune in inglese: la /t/ in *missed seven* [mist sevn] è marcata principalmente da una riduzione nell'ampiezza spettrale e da alcune variazioni della porzione di rumore continuo. Da questo contesto sono assenti i parametri classici associati al tratto [+ consonantico] e [- continuo]: netta differenza tra l'involuppo spettrale del *burst* e delle porzioni di segnale che lo seguono (Stevens, 1989, 2002; Stevens & Blumstein, 1978), cambiamento repentino dell'ampiezza del segnale al rilascio della consonante. Tuttavia, le variazioni della forma dello spettro ai confini tra le sezioni ad alta e bassa ampiezza sono sufficienti perché la sequenza sonora venga percepita da parlanti inglesi nativi come un suono [+ consonantico -continuo +anteriore +coronale] preceduto e seguito da segmenti [+ consonantico +continuo +anteriore +coronale].

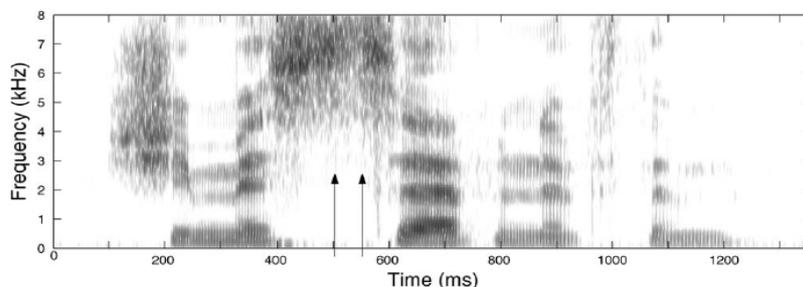


Figura 3: Sonagramma della sequenza *She missed seven of them*

Come accade visivamente per i Triangoli di Kanizsa, quando il contesto acustico è coerente, la porzione di suono viene elaborata attraverso un processo di *creazione* di un confine che viene *classificato* grazie alle proprietà degli elementi del contesto. La differenza tra prominente uditiva di un passaggio acustico e localizzazione di confini apparenti nei Triangoli di Kanizsa è che la prima mette a fuoco *ciò che è nel segnale*, mentre la seconda deriva dalla sovrapposizione di una proprietà che *non è nel segnale*.

L'illusione ottica di Müller-Lyer, invece, riguarda modifiche percettive di dimensione degli oggetti rappresentati: in fig. 4, delle due frecce esattamente della stessa lunghezza, ma

con parti terminali differenti: si tenderà a percepire più lunga quella con i prolungamenti della freccia rivolti verso l'esterno.

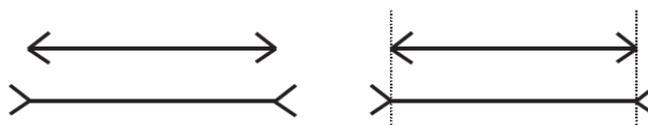


Figura 4: Illusione di Müller-Lyer

Il parallelo a livello uditivo è costituito dagli esperimenti di Carlyon et alii (in stampa): tra due porzioni di rumore confinanti con differente larghezza di banda [fig. 5a], la presenza di rumore a banda larga (WBN) aumenta la durata percepita del rumore a banda stretta (NBN), come verificato dagli aggiustamenti effettuati dagli ascoltatori (v. fig. 5b).

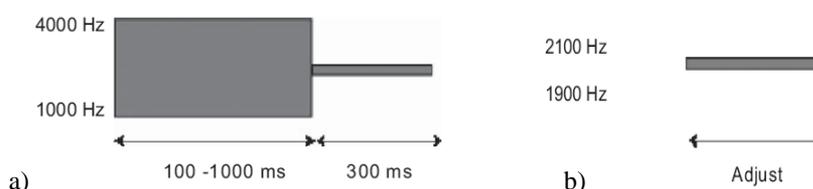


Figura 5: a) sequenza di rumore a banda larga e rumore a banda stretta (WBN+NBN); b) rumore a banda stretta in isolamento (Carlyon et alii, in stampa)

I due esperimenti sottolineano il ruolo che, nel sistema percettivo, assumono confini bruschi vs. gradualmente tra uno stato e l'altro e possono avere rilevanza anche per la percezione fonetica.

Gli esempi in fig. 1b (*let* vs *led*) evidenziano un confine diverso tra segmento sonorante (vocalico) e segmento fonologicamente [+sonoro] e [-sonoro] (con un più netto *offset* dei parametri nel secondo caso). Risultati simili sono riportati nei lavori di Diehl et alii, (2004) che hanno investigato la salienza percettiva di questi contrasti per le consonanti occlusive. In stimoli sintetici che simulano sequenze V-C(+ostruente), quando i valori frequenziali di F1 ed  $f_0$  della vocale al confine con il segmento consonantico sono più bassi che nel resto della vocale, di norma si percepisce una consonante con tratto [+sonoro]. Nel parlato spontaneo si registra una simile diminuzione; inoltre i tratti vocalici, di norma, tendono a permanere all'interno della fase di chiusura della consonante. Al contrario, i valori di F1 della vocale rimangono alti quando la consonante è [-sonora] e si verifica successivamente una brusca interruzione di tutti i parametri; ne risulta una fase di silenzio piuttosto prolungata e quindi un contrasto acustico maggiore.

Se si assume, quindi, che la NBN funzioni come periodo di chiusura di un'occlusiva e la WBN come il segmento sonoro precedente (una vocale), la sequenza WBN+NBN può essere assimilata ad una sequenza VC nella quale il confine tra V e C è brusco, quando C è caratterizzato dal tratto [-sonoro]. In sostanza la sequenza WBN+NBN verrebbe percepita *olisticamente* piuttosto che nelle sue parti isolate; questo pattern acustico suscita la risposta uditiva ad una differenza fisica che viene sfruttata e soggiace alla distinzione fonologica tra segmenti [+sonori] e [-sonori]. E' molto difficile, dunque, distinguere tra lo stimolo e il suo contesto: la sequenza VC# nel parlato può essere percepita altrettanto olisticamente.

## 5. ILLUSIONI DI CONTINUITÀ

Nelle illusioni di continuità sensazioni oggettivamente discontinue vengono percepite come un percolato completo grazie al contesto nel quale occorrono. Il contesto permette di “riempire” la sensazione mancante, e percepire qualcosa che non è oggettivamente presente. Nel quotidiano, illusioni di continuità si hanno in situazioni in cui soltanto una parte di un oggetto è percepibile, ma il contesto dove quella parte è calata indica che il resto dell’oggetto è in qualche modo nascosto. E’ ragionevole supporre che molte illusioni di continuità nascano dall’interazione tra conoscenza e sensazione<sup>2</sup>.

Esistono molte illusioni di continuità di tipo visivo. Una delle più divertenti e ‘convincenti’, è quella discussa da Bregman (1981, 1990; v. fig. 6) nel contesto della sua teoria dell’*Auditory Scene Analysis*.

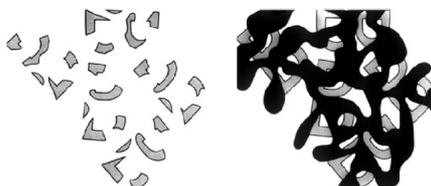


Figura 6: l’illusione delle “B” (<http://www.michaelbach.de/ot/>)

Mentre non si riesce a vedere qualcosa di sistematico nel pannello di sinistra, in quello di destra appare un insieme di *B* casualmente disposte: la macchia d’inchiostro, apparentemente ‘casuale’, fa emergere l’informazione sistematica.

Già i primi esperimenti fonetici effettuati attraverso la manipolazione di parametri acustici avevano dimostrato che l’identità fonemica può essere veicolata oltre che da proprietà fisiche dello stimolo in sé, anche da proprietà del contesto immediato o a lungo termine (per una rassegna cfr. Hawkins, 2004).

Un parallelo con gli esperimenti discussi da Bregman è la dimostrazione di Fowler e Smith (1986), ottenuta utilizzando tecniche di ‘*cross-splicing*’ di *schwa* in sequenze come /ibəbi/ e /abəba/: quando le porzioni periodiche corrispondenti agli *schwa* vengono invertite nei rispettivi contesti, gli inglesi nativi percepiscono approssimativamente [ibəbi] e [abəba], mentre, nei loro contesti originali, percepiscono i due suoni come simili. Il contesto inappropriato mette dunque in evidenza a livello percettivo una differenza (dovuta a coarticolazione *vowel-to-vowel*) esistente nello stimolo, ma normalmente trascurata. In inglese, per essere percepito come *schwa*, è sufficiente che un suono sia più breve e centralizzato rispetto alle vocali adiacenti: i tratti di altezza e centralità ([-alto, -basso, -anteriore, -arretrato] ) emergono tuttavia, piuttosto che dalle proprietà fisiche del suono, grazie al contesto che, come la macchia d’inchiostro di Bregman, può essere considerato come ciò che elicitava l’informazione sistematica.

Un’altra illusione ottica di continuità è l’*illusione di Poggendorff*, (fig. 7). Qui, l’aggiunta del rettangolo solido alle linee disposte nel pannello a sinistra dà l’impressione

---

<sup>2</sup> L’illusione di continuità non deve essere confusa con comportamenti di deduzione logica, con i quali ha qualcosa in comune: perché si possa parlare di ‘illusione’, la riproduzione del percolato dev’essere difficile, se non impossibile, da effettuare in altro modo.

di due linee parallele, una solida e l'altra punteggiata, laddove ce ne sono tre, due corte e una lunga, esattamente come nel pannello di destra.

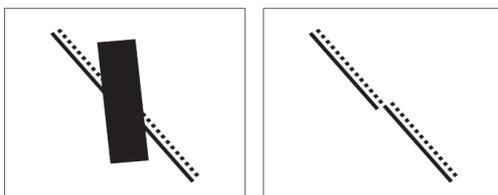


Figura 7: una versione dell'illusione di Poggendorff

Sono illusioni di continuità, che si verificano sull'asse temporale, anche gli effetti del *tracking motion* (citati da Bregman, 1990, p.41, come esempi di movimento illusorio).

Analogie uditive del movimento illusorio e dell'illusione di Poggendorff sono costituite dagli esperimenti di mascheramento di toni assoluti o costanti attraverso bande di rumore (Ciocca & Bregman, 1987). Il parametro manipolato è di solito la relazione tra le frequenze contigue al rumore a banda larga, che produce una sensazione di continuità o cambiamento di traiettoria: l'illusione dev'essere compatibile con il tono in entrata e in uscita dalla banda di rumore in maniera sistematica e predicibile. Le somiglianze nella frequenza fondamentale, la distribuzione spettrale e la localizzazione apparente del suono nello spazio sono tutti fattori importanti per l'illusione. Inoltre è più probabile che un'illusione di continuità sia efficace se è conforme ad un fenomeno del mondo reale che coincida con uno schema mentale, o con la conoscenza di un *pattern* preesistente.

Quasi ogni fenomeno conosciuto in acustica e percezione del parlato è pertinente per questo tipo di illusione, soprattutto se si guarda all'ampia letteratura sul ruolo delle transizioni formantiche per la percezione del luogo di articolazione. Anche gli esperimenti sull'intelligibilità sia del parlato naturale che artificiale in condizioni di ascolto difficili dimostrano che la continuità del pitch, l'ampiezza, la struttura spettrale etc. sono cruciali per la comprensione del parlato.

Gli studi di Carlyon et alii (2002) mostrano che quando F1 e F2 di due vocali sintetiche sono alternativamente intermittenti, il risultato non è percepito come vocale; invece quando il silenzio tra ogni formante intermittente è sostituito da un rumore la cui larghezza di banda, durata e intensità siano tali da mascherare le formanti, viene percepita una vocale illusoria. Henrich et alii (2008) attraverso analisi di tipo fMRI dimostrano che lo stimolo illusorio produce un'attivazione delle aree linguistiche del cervello sensibili al parlato. I modelli di percezione uditiva che si fondano sulla costruzione di un percetto continuo da un'informazione intermittente (e.g. Cooke, 2006; Cooke & Ellis, 2001) sono parzialmente motivati da dati di questo tipo.

Le numerose analogie dell'illusione di continuità con il parlato includono il processo di "ricostruzione spettrale" e gli effetti della "ricostruzione fonemica" discussi da Warren (1999). La "ricostruzione fonemica" occorre quando gli ascoltatori non riescono a distinguere tra una parola integra ed una nella quale un suono è stato sostituito da rumore, ad esempio la [s] di *legislature*. La "ricostruzione spettrale" si riferisce ad effetti percettivi simili in condizioni in cui le parti del segnale sono presenti su tutti gli intervalli di tempo, ma parti dello spettro della frequenza sono sostituite da rumore. In questo tipo di esperimenti alcuni parametri, come le relazioni tra ampiezza e frequenza, sono critici e

devono avere una qualche plausibilità. Warren (1999, p. 136) riporta che si raggiunge il massimo dell'intelligibilità, durante una ricostruzione spettrale, quando i livelli di rumore sono inferiori di circa 10dB rispetto al livello del segnale di picco. Viceversa, si ha ricostruzione fonemica quando i livelli di rumore sono abbastanza alti da mascherare il segnale<sup>3</sup>. Samuel (1981, 1987) evidenzia che la ricostruzione è più robusta quando il "fonema ricostruito" è sostituito da un suono che si prevede possa rimpiazzarlo- un rumore bianco è un *masker* migliore di fricative e occlusive più che di sonoranti. La conoscenza circa le parole possibili (*'lexical uniqueness'*) influenza le decisioni e le risposte, mentre è noto che l'attivazione lessicale può, da sola, indurre forti percetti fonemici, ad ulteriore prova del ruolo del contesto sulla percezione di continuità (Samuel, 1997).

L'interpretazione della ricostruzione fonemica come tipo di illusione di continuità è possibile soltanto se si assume un'unità più grande del fonema o del fono (cfr. Bashford et alii, 1988). Ciò che è cruciale è quindi il pattern uditivo della parola: il fenomeno potrebbe quindi essere battezzato "*illusory word completeness*", postulando uno statuto di categoria auto-organizzata per l'unità lessicale (almeno per quelle lingue per le quali la nozione di parola è non problematica<sup>4</sup>).

## 6. ILLUSIONE ED ASPETTATIVA

In certe particolari condizioni il nostro cervello fa emergere straordinarie illusioni. Tra queste sono note, ad esempio, quella della *maschera di Hor*<sup>5</sup> e quella stupefacente della *maschera di Chaplin*: una maschera di Charlie Chaplin in versione *cartoon* ruota lentamente sull'apice di un bastoncino. Il davanti della maschera è convenzionalmente convesso, ed il rovescio vuoto, o concavo. Quando in un primo momento appare il rovescio della maschera, sembra concavo (come realmente è). Ma non appena appaiono i caratteri "fisici" della maschera (gli zigomi e le orbite degli occhi), l'immagine ruota in modo tale che il viso sembri convesso; il naso è proiettato all'infuori ed il viso sembra ruotare nella direzione opposta<sup>6</sup>. Per dirla con le parole dello stesso Gregory (2005): " il cervello si rifiuta di vedere il rovescio della maschera come vuoto perché è un fenomeno molto improbabile, e questo dimostra l'immenso potere della conoscenza di tipo *top-down*, che [...] forza un'illusione straordinaria, nella quale l'informazione sensoriale è cancellata dalla conoscenza derivante dal passato, nel quale tutte le facce che si sono viste avevano il naso proiettato verso l'esterno". La conoscenza/esperienza a cui Gregory si riferisce si contrappone non soltanto alla sensazione visiva in sé, ma alla stessa consapevolezza di vedere ruotare una maschera vuota. L'argomentazione dell'autore è che l'illusione "puramente percettiva" è mediata da un flusso ventrale di informazioni che procede lungo i lobi temporali, mentre i processi cerebrali *goal-directed* di azione-percezione sono mediati attraverso un flusso dorsale che passa attraverso i lobi parietali e sembrano essere meno suscettibili all'influenza dell'illusione.

---

<sup>3</sup> Analogamente, è possibile avere esperienza dell'illusione di Poggendorff soltanto quando il rettangolo oscurante è opaco.

<sup>4</sup> Ovviamente le lingue indoeuropee. Per tipi linguistici diversi (ad esempio le lingue Athabaskan) si veda McDonough, 2001.

<sup>5</sup> [Http://dragon.uml.edu/psych/hor.html](http://dragon.uml.edu/psych/hor.html)

<sup>6</sup> Un'idea migliore si può avere visitando il sito di Richard Gregory all'URL: <http://www.richardgregory.org/experiments/index.htm>

Effetti simili occorrono a livello uditivo, per esempio nei noti esperimenti sull'effetto *McGurk*<sup>7</sup> (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). La dimostrazione originale mostra in video un parlante inglese che articola la sequenza /gaga/, mentre contemporaneamente si ascolta la sequenza /baba/. Il 98% dei partecipanti all'esperimento ha risposto che la sequenza articolata era /dada/, e nessuno stimolo fisico riflette questa risposta, che è un'effettiva "fusione" delle due sequenze proposte. Quando le due stringhe sono state invertite (in video /baba/ e in audio /gaga/) le risposte sono state più variegate: 31% /baba/, 11% /gaga/ e il 54% risposte "combinare", come /gabga/ o /gaba/. Questa illusione dipende principalmente da un delicato bilanciamento tra conoscenza e sensazione, insieme ad una parte di informazione "migliore" o più "affidabile" dello stimolo: vedere labbra che si chiudono è certamente indice di labialità; non vederle chiudersi certamente è indice di non labialità, ma non è un indicatore sufficiente per l'individuazione certa di altri luoghi di articolazione. La percezione finale, quindi, dipende maggiormente dallo stimolo uditivo. Un'occlusiva labiale manca di una significativa cavità frontale, quindi ha uno spettro del *burst* più 'piatto' (*flat*) che spesso è anche più basso in ampiezza. In contrasto, gli spettri del *burst* di occlusive non labiali presentano forme più distinte, dovute alle risonanze della cavità frontale, e sono spesso più alte in ampiezza relativa perché la costrizione ha un rilascio più rapido. Le risposte in combinazione si hanno quando entrambi gli stimoli non risultano ambigui in nessuna delle due modalità, ad esempio quando lo stimolo visivo /baba/ è accoppiato allo stimolo altamente distintivo /gaga/. Ma quando le labbra non si sono chiaramente chiuse, un /baba/ udito sarà interpretato come un'occlusiva non labiale non ben articolata, coperta da altri suoni e così via. La ragione per la quale, invece, la risposta agli stimoli visivi /baba/ e uditivo /gaga/ sia /dada/ (e non altre possibilità) è che quando lo stimolo uditivo fisico presenta uno spettro del *burst* più 'piatto' indicativo di labialità, ma la vista chiaramente mostra che le labbra non si sono chiuse, il sistema percettivo "assume" che un qualche rumore di banda larga ad alta frequenza abbia nascosto le proprietà spettrali distintive tipiche per il *burst* dell'occlusione in /da/. E' possibile anche che la risposta sia influenzata dall'alta frequenza di alveolari, in inglese come in altre lingue (Denes, 1963; Sekiyama & Tohkura, 1991; Studdert-Kennedy & Shankweiler, 1970): /d/ corrisponderebbe meglio alle informazioni disponibili e verrebbe dunque percepito attraverso normali elaborazioni probabilistiche.

Altre ricerche successive hanno confermato che l'effetto McGurk è suscettibile ad una grossa somma di influenze, varia con il variare di proprietà idiosincratice dello stimolo (Green et alii, 1988; Fixmer & Hawkins, 1998), del parlante stesso (Munhall et alii, 1996; Gagné et alii, 2002), della lingua (Sekiyama, 1996; Sekiyama & Tohkura, 1991, 1993; Iverson et alii, 1998; Sams et alii, 1998) e sono chiaramente mediate dalla conoscenza.

Concludendo, si può arguire che i processi percettivi abitualmente "soppesano" le informazioni sensoriali uni- e multi-modalità per arrivare al percepito più probabile.

## 7. ILLUSIONE ED ATTENZIONE

Sembra generalmente accettato che l'attenzione guidi la selezione degli oggetti per elaborazioni successive. Esperimenti sull'attenzione spaziale (Hawkins et alii, 1990) suggeriscono che l'attenzione migliora la sensibilità percettiva.

---

<sup>7</sup> sono facilmente rintracciabili in rete video dimostrativi dell'effetto McGurk: una dimostrazione affidabile è all'URL: <http://youtube.com/watch?v=aFPtc8BVdJK>.

L'attenzione può essere molto focalizzata su cose particolari o può essere relativamente defocalizzata. Nel caso di input altamente strutturati e complessi come il parlato, i cambi di focus dell'attenzione sono molto comuni e fortemente *task-dependent* (Mattys et alii, 2005; Werker & Tess, 1984). L'abilità di "stringere" il focus dell'attenzione probabilmente rafforza il processo di "addestramento percettivo" dell'ascoltatore rendendo la percezione altamente adattiva. E' stato dimostrato (Seitz et alii, 2005) che un addestramento percettivo volto ad un segnale visivo in movimento può, nelle giuste circostanze, indurre la percezione di un movimento illusorio. E' presumibile che gli stessi processi influenzino l'addestramento percettivo nel dominio uditivo, che è necessariamente di tipo dinamico.

I modelli che fanno capo alla *Adaptive Resonance Theory* (ART; Carpenter & Grossberg, 2003; Grossberg, 2005) offrono una rappresentazione per le variazioni nell'attenzione compatibile con l'idea che le unità linguistiche comprendano un aspetto illusorio<sup>8</sup>. L'assunto primario dell'ART è che l'aspettativa è collegata all'input sensoriale. Quando la sensazione in entrata corrisponde all'aspettativa, ne consegue una "risonanza", equivalente ad una identificazione positiva di un'unità rilevante; è il percolato che vince su tutti gli altri concorrenti e raggiunge lo stadio di consapevolezza, anche se solo momentaneamente. Più specificamente, la risonanza è un *loop* di *feedback* positivo tra proprietà sensoriali del segnale e *item* prefissati nella memoria attiva, identificato con un certo grado di sicurezza. Questo *loop* è fortemente influenzato dalla conoscenza o dai "primitivi", cioè dalle aspettative di una risposta possibile, presenti nella memoria a lungo termine. Quando c'è risonanza, si sviluppa attraverso la memoria attiva un'onda di attività risonante che lega la percezione di unità linguistiche più piccole in unità più grandi e le fa emergere alla percezione conscia dell'ascoltatore. Una risonanza quindi collega la sensazione, il contesto antecedente e l'aspettativa; di conseguenza, unità di dimensione maggiore sono più accessibili all'identificazione conscia rispetto alle loro componenti.

Il parametro ART di "vigilanza" specifica la frazione minima dell'input sensoriale presente nell'aspettativa (*pattern* di tipo *top-down*), necessaria perché si crei una relazione tra sensazione ed aspettativa e quindi l'inizio di una risonanza. Una bassa vigilanza porta ad ampie generalizzazioni, e una vigilanza più alta a categorizzazioni più dettagliate e definite. Il livello più alto di vigilanza è equivalente all'apprendimento esemplare. Questo tipo di modello rende conto dei processi di apprendimento veloce (*fast learning*) senza perdere la memoria di elementi appresi precedentemente, e di riconoscimento di fenomeni inusuali.

Nella teoria ART le illusioni percettive hanno un ruolo importante (Grossberg, 2005; Pinna & Grossberg, 2006). In sostanza, gli assunti generali della ART e della *Auditory Speech Analysis* sono molto simili e tra loro collegati: in entrambi gli approcci, è l'interrelazione tra sensazione e aspettativa che causa la percezione della 'realtà', indipendentemente dalla natura fisica o illusoria del percolato.

Possiamo legare queste osservazioni alla teoria dei tratti distintivi in questo modo. In primo luogo, tutto o quasi il segnale acustico interpretabile include proprietà che possono essere identificate con un alto grado di sicurezza, verso le quali è probabile che l'attenzione, anche se solo momentaneamente, sia direzionata. Queste proprietà ad alta affidabilità funzionano come "ancore" (*anchor points*) attorno alle quali si costruisce il resto del percolato; l'interpretazione percettiva degli *anchor points* varierà relativamente

---

<sup>8</sup> Per una rassegna cfr. Grossberg 2003; Hawkins & Smith 2001.

poco mentre si compongono, all'interno del percolato costruito, le parti più ambigue del segnale. Per il parlato, i correlati acustici di tratti distintivi rappresentano alcune di queste proprietà: ad esempio, il tratto [+ sonoro] è non ambiguo in molti segmenti [+ sonoranti] delle lingue (v. §2.1). Il tratto potrebbe non essere cruciale per l'identificazione dell'esatta qualità delle sonoranti, ma potrebbe essere fondamentale per la percezione della struttura ritmica dell'enunciato; è risaputo infatti che il ritmo è un parametro centrale per l'intelligibilità linguistica, quindi si suppone che i correlati acustici di [ $\pm$  sonoro], che si relazionano con la sillabificazione abbiano una profonda influenza a livello percettivo.

In secondo luogo, sebbene l'attenzione dell'ascoltatore sia focalizzata per la maggior parte su unità linguistiche di ampie dimensioni (come parole o sintagmi), talvolta essa può essere diretta consciamente ai tratti distintivi, per aumentare il grado di vigilanza (come predice l'ART): l'ascoltatore attento riesce a spostare rapidamente l'attenzione tra livelli di informazione diversi, vale a dire tra livelli diversi della struttura linguistica.

Inoltre, gli stessi tratti distintivi possono essere illusori. In una situazione di ascolto normale, nella quale il segnale ha un buon rapporto di coincidenza con le aspettative dell'ascoltatore, è sufficiente la presenza nel segnale di proprietà con basso livello di affidabilità per innescare la ricostruzione del percolato. Quando nel segnale mancano le proprietà attese o esistono delle aspettative in competizione, il livello di vigilanza cambia e l'attenzione si concentra sui dettagli del segnale.

Infine, una volta innescato il processo illusorio, la sensazione reale lascia il posto alla memoria della sensazione illusoria. Ne è prova l'esperienza comune per cui spesso non ricordiamo le parole esatte di un interlocutore, ma piuttosto il senso generale che ne abbiamo derivato.

## 8. LA PERCEZIONE DI “OGGETTI Uditivi”

### 8.1. Attivazione distribuita, plasticità, multi modalit  della percezione

Non esiste una definizione chiara per il concetto di “oggetto uditivo” e in ambito linguistico il problema   pi  complesso in quanto esistono moltissime unit  di percezione potenziali; inoltre, come discusso in §7, l'attenzione   mediata dal significato, che a sua volta presenta una relazione a tutt'oggi ancora poco compresa con il segnale fisico.

Nell'ipotesi qui avanzata che i processi di percezione, sia uditiva che visiva, siano governati da meccanismi equivalenti, il termine ‘oggetto uditivo’ si pu  considerare preliminarmente *una metafora che descrive un'astrazione costruita sulla base di un'attivazione percettiva multisensoriale*: un debole oggetto uditivo pu  derivare anche da stimoli di tipo visivo, come quando “si ascoltano” le parole che si leggono. Secondo questo punto di vista, il percolato   reale, in senso neurochimico, nel momento in cui sensazione e memoria si sovrappongono, producendo la percezione di un oggetto reale.

Questo processo costruttivo appartiene fondamentalmente ad elaborazioni di tipo corticale, come mostrano alcuni esperimenti di illusione ottica sull'effetto di *grating* (Meng et alii, 2005) e sembra essere tipico anche della percezione uditiva (Bregman, 1990; Deutsch, 1999, 2007; Warren, 1999):   ragionevole supporre infatti che gli stimoli uditivi attivino domini sensoriali multipli, e che si manifesti quindi una ricca attivit  di tipo corticale. La maggior parte delle situazioni in cui il parlato viene percepito contempla il contatto visivo o la memoria visiva dell'interlocutore. Inoltre, la maggioranza degli ascoltatori ha grande esperienza nell'associare i suoni linguistici a *feedback* di tipo cinetico e cinematografico del tratto vocale, cos  come possiede informazioni tattili relative al tratto

vocale stesso. Quindi, la situazione *default* di ascolto è generalmente multimodale, almeno una volta che il segnale ha raggiunto la corteccia cerebrale.

E' dimostrato inoltre che l'integrazione multisensoriale nelle fasi iniziali dell'elaborazione è la norma (Ghazanafar & Schroeder, 2006; Murray et alii, 2005). Tutta l'elaborazione percettiva è inoltre massicciamente distribuita sia lungo la corteccia cerebrale che nelle regioni corticali, tradizionalmente considerate unisensoriali, e questo fa in modo che il percepito sembri 'più di una somma delle sue parti' (Adams & Janata, 2002; Murray et alii, 2005; Sussman et alii, 2002; Weinberger, 2004) e processi di tipo cognitivo coinvolgono buona parte del cervello in elaborazioni altamente distribuite ed interattive. Le aree tradizionalmente associate all'elaborazione del parlato sono distribuite, presentano una certa plasticità e sottintendono elaborazioni sintetiche, multimodali o amodali. E' significativo, per esempio, che la percezione dei movimenti della bocca venga elaborata nelle stesse regioni che si attivano durante la percezione uditiva (l'area di Broca e la parte posteriore del solco temporale superiore; Pelfrey et alii, 2005; Santi et alii, 2003) e che la lingua dei segni in persone non udenti venga elaborata nelle stesse regioni di *processing* del parlato in soggetti udenti (Campbell et alii, 2008).

### 8.2. Il sistema corticofugale: una base per l'interazione tra conoscenza e sensazione

Il ruolo dei processi di tipo *top-down* non è in discussione. Meno note, tuttavia, sono le prove di tipo neuroanatomico e neurofisiologico che dimostrano come questi processi si ripercuotano sulle primissime fasi dell'elaborazione di tipo uditivo. Recenti verifiche empiriche lasciano supporre che il *feedback* proveniente dalle aree di livello superiore influenzi le fasi iniziali dell'elaborazione, poiché la corteccia uditiva primaria imprime nella memoria tracce del comportamento di suoni specifici (cfr. Fritz et alii, 2003; Fritz et alii, 2005; Weinberger, 2004). Quindi alla base della plasticità della percezione indotta dal contesto c'è presumibilmente il sistema corticofugale, che incanala le informazioni uditive attraverso le aree cerebrali (Khalifa et alii, 2001; Weinberger, 2004; Winer & Lee, 2007). Più in generale, il sistema corticofugale può essere visto come sede di calibrazione o mediazione tra sensazione uditiva in entrata e decisioni, in un modo più o meno simile a quanto previsto dal modello percettivo di *Analysis-by-Synthesis* di Halle & Stevens (1962). L'importanza di tutto ciò per la costruzione degli oggetti uditivi è scontata.

E' largamente provato che l'elaborazione nella corteccia celebrale coinvolge complesse operazioni di collegamento, strutturazione ed archiviazione dell'informazione.

La corteccia uditiva primaria è la prima area della corteccia cerebrale a ricevere lo stimolo uditivo; essa risponde in maniera distinta ai diversi stimoli acustici complessi che vengono collegati, nella stessa corteccia e nelle aree adiacenti, creando 'oggetti uditivi', che verrebbero interpretati come correlati acustici di tratti distintivi (Griffiths & Warren, 2002, 2004; Jacquemot et alii, 2003; Näätänen et alii, 2001; Obleser et alii, 2006; Nelken et alii, 2003; Scott, 2005) ed elaborati come tali nelle zone cerebrali iniziali del ciclo uditivo, se non ancora prima. Tuttavia altre prove dimostrano che l'elaborazione in categorie fonetiche di uno stimolo uditivo coerente (di tipo *speech-like* e *non-speech-like*) ha luogo in diverse parti dell'intera corteccia (Benson et alii, 2006; Binder et alii, 2004; Raizada & Poldrack, 2007), coinvolgendo anche le strutture non corticali (il sistema limbico, il talamo, i gangli basali e il cervelletto) incluse nel sistema corticofugale (Winer, 2006).

Insomma, è possibile concludere che gran parte dell'elaborazione cerebrale del parlato è di tipo integrato, altamente astratto e contemporaneamente legato alla sensazione e contestualizzato, esattamente come per l'elaborazione visiva (Bar & Aminoff, 2003).

### 8.3. *Oggetti uditivi e unità linguistiche*

Il termine ‘oggetto uditivo’, dunque, poggia sull’idea che la corteccia colleghi le informazioni provenienti dal sistema sensoriale a quelle residenti in memoria in unità coerenti. Griffiths & Warren (2004) suggeriscono che un oggetto uditivo è una rappresentazione spettrotemporale familiare (sonogramma tridimensionale – frequenza, ampiezza e tempo – trasformato in *pattern* di attivazione spettrotemporale attraverso processi neurali), ed accentuano l’importanza di un *pattern* coerente e della natura astratta del percolato. Gli autori osservano che le illusioni uditive e oggetti uditivi sono compatibili, e che gli oggetti uditivi possono avere una dimensione temporale, come si sostiene anche in questo lavoro. Tuttavia il presente articolo differisce dalla posizione di Griffiths & Warren nell’ipotesi secondo la quale le unità linguistiche possono formare oggetti uditivi.

Griffiths & Warren (2004, p. 887) notano che “i meccanismi per l’estrazione iniziale di tratti uditivi elementari sono cruciali per l’astrazione delle proprietà oggettive di ordine più alto”, ma non focalizzano l’attenzione sull’assunto iniziale di questo articolo: le stesse proprietà linguistiche rilevanti (tratti distintivi, fonemi, sillabe) sono spesso difficili da individuare nel segnale. Gli autori, inoltre, discutono di tratti uditivi in senso generale, intendendo qualsiasi proprietà acustica spettrotemporale di un dato segnale; per la percezione del parlato assumono il punto di vista standard secondo cui il significato viene compreso solo dopo aver identificato una serie di schemi formali (presumibilmente forme fonologiche, parole, etc.).

Di contro, questo articolo propone tutte le unità linguistiche come potenziali candidate, considerandole categorie emergenti (o astrazioni) sintetizzate attraverso un processo bayesiano derivante da una combinazione del segnale e della conoscenza/memoria: gli ascoltatori, intenti a comprendere il significato, utilizzano il significato stesso per costruire la forma, ugualmente o più di quanto facciano l’inverso. In altre parole, nella maggior parte della comunicazione quotidiana, la forma è utile agli ascoltatori fin quando li aiuta a capire il significato. La sensazione che induce l’identificazione di un oggetto uditivo è in genere un *pattern uditivo*, che può comprendere proprietà uditive più dettagliate e raffinate rispetto a quelle che inducono identificazione fonemica o allofonica, e/o proprietà meno sottili, e può essere significativamente maggiore o minore del fonema. Il ‘parlato non accurato’ della comunicazione quotidiana dimostra come il significato e il contesto dell’interazione possano essere usati per costruire rappresentazioni mentali della forma (v. §2).

Inoltre, mentre la ricerca neuroscientifica ritiene, implicitamente o esplicitamente, che la percezione del parlato avviene in una serie di fasi rigidamente separate, dove le unità fonologiche sono identificate prima delle parole ed il significato viene identificato alla fine (Binder et alii, 2004; Griffiths & Warren, 2004), di contro qui si argomenta che l’identificazione di *pattern* e quindi di oggetti uditivi che rappresentano unità linguistiche non segue necessariamente un rigido ordine sequenziale. Le unità linguistiche di livello superiore possono essere identificate prima o contemporaneamente alle unità di livello inferiore, come accade con in confini prosodici (Fougeron & Keating, 1997; cfr. anche Hawkins, 2003; Hawkins, in press; Hawkins & Local, 2007; Hawkins & Smith, 2001).

Infine, Griffiths & Warren (2004) ignorano dettagli acustico-fonetici che potrebbero ridurre i loro dubbi riguardo ai paralleli tra illusione uditiva e visiva. Ad esempio, essi pongono l’accento sull’importanza di confini e contorni, ed evidenziano le difficoltà di applicazione di questo concetto al segnale acustico. Nel §4.1 si è argomentato che certi contrasti tra confini netti vs. gradualmente sono attributi della percezione fonetica fondamentali e

particolarmente rilevanti nell'identificazione di tratti fonologici distintivi, come discusso da Stevens (2002) e Fischbach et alii (2001).

Per concludere, questo lavoro adotta un punto di vista in qualche modo più ricco sulla sostanza di un oggetto uditivo, principalmente perché tende a considerare il segnale come veicolo di informazioni linguistiche e comunicative di ogni livello piuttosto che concentrarsi sul solo contrasto fonologico.

Prove a supporto della possibilità di considerare le unità linguistiche come oggetti uditivi derivano dal fenomeno del *'pop-up'*, termine che descrive l'esperienza di percepire chiaramente qualcosa che già si conosce in di un contesto fortemente distorto. Il *pop-up* occorre nella vista, nell'udito e presumibilmente anche in altre modalità sensoriali. Nella modalità uditiva, il *pop-up* è alla base dei fenomeni ben conosciuti che governano l'identificazione di parole e sillabe in contesto di rumore (Kalikow et alii, 1977; Miller et alii, 1951). Nonostante sia difficile differenziare tra il riconoscimento di una forma fonologica e l'effetto di un'informazione memorizzata, è ovvio che l'esperienza influisce con la comprensione interagendo con il livello del significato. Queste argomentazioni suggeriscono che qualsiasi unità linguistica funzionalmente rilevante per un ascoltatore può diventare un oggetto uditivo. Per certi versi, a causa della loro indipendenza dal significato, i tratti fonologici distintivi possono sembrare i candidati meno idonei per la funzione di oggetto uditivo. Ma la loro natura formalmente astratta, la loro mescolanza di proprietà acustico-uditive generiche accoppiata alle loro specificità linguistiche e l'influenza del contesto che ne permette l'identificazione in ogni circostanza, li rendono non dissimili da qualsiasi altra unità linguistica derivabile da un segnale, e quindi candidati ragionevoli a svolgere la funzione di oggetti uditivi.

## 9. CONCLUSIONI

Si è argomentato che gli oggetti uditivi, se esistono, probabilmente coinvolgono molte zone del cervello, implicano analisi di tipo generale e dettagliato, e hanno il loro fondamento in elaborazioni basiche della memoria e della percezione, modulate dall'attenzione, che a sua volta influenza sia la sensibilità percettiva che i *bias*. Il *matching* tra memoria e sensazione comincia contemporaneamente alla sensazione, con il percolato reale che riflette un delicato equilibrio tra aspettativa e sensazione, come indicato da una serie di illusioni dimostrabili. Questi processi, ampiamente discussi in letteratura e variamente formalizzati (Cooke, 2006; Greenberg, 2006; Husain et alii, 2004; Jazayeri & Movshon, 2007; Klatt, 1979; Nguyen et alii, in press; Tuller 2003; Warren, 1999), consistono in una complessa corrispondenza di indici ricostruita nella memoria, indicata spesso dalla nozione di *template*. I *templates* potrebbero essere non più che *pattern* profondamente appresi attraverso l'esperienza: la risposta finale è il risultato di complesse strategie di decodifica (a loro volta influenzate dall'attenzione e dall'aspettativa) che soppesano le proprietà intrinseche dello stimolo, il contesto, e la possibilità di 'riempire' un'unità percettiva significativa. Per tutti gli ambiti o quasi, questi *templates* sono presumibilmente multimodali, e saranno multidimensionali per le unità linguistiche.

La posizione qui espressa non è lontana dalle premesse fondamentali della *Quantal Theory* e della teoria dell'invarianza uditiva. La *Quantal Theory* incorpora esplicitamente l'idea che l'elaborazione uditiva possa creare distinzioni linguistiche anche quando le differenze acustiche siano poche; afferma che questi processi sono automatici e di basso livello; postula la non linearità del sistema uditivo, compatibile con il tipo di influenza *top*

*down* enfatizzata in questo articolo; afferma anche che l'esperienza può modificare alcune di esse. Una forma estrema di teoria dell'invarianza uditiva risulterebbe invece incompatibile con questa posizione. Si può tuttavia valutare la nozione di 'invarianze relazionali' nei termini di fattori descrittivi delle manifestazioni fisiche di particolari tratti distintivi: come discusso in §7, si suggerisce che molte invarianze acustiche relazionali, nonostante non siano affidabili al 100% in tutti gli stili di parlato e in tutti i contesti, forniscono indici di alta certezza di un contrasto particolare. Quindi, quando presenti, esse funzionano da *anchor points* o 'isole di affidabilità' nell'input sensoriale, attorno alle quali viene organizzata l'informazione rimanente.

Il punto di vista qui espresso è compatibile anche con un certo numero di modelli psicolinguistici di riconoscimento di parole (Elman, 2004; Gaskell & Marslen-Wilson, 2002; Tabor et alii, 1997). E' necessario ulteriore lavoro sul parlato che provi a comprendere come le proprietà spettrottemporali si raggruppano in oggetti uditivi

L'enfasi sull'esperienza consente di far emergere effetti gradienti e differenze individuali o differenze tra diverse categorie linguistiche (Wurm, 1997) e tra diverse comunità linguistiche. Non tutte le persone sperimentano, infatti, esattamente la stessa cosa. Gli artisti potrebbero non essere ingannati dall'illusione di Müller-Lyer e da quella di Poggendorff, e neanche da quelle che includono giochi di prospettive, mentre altre persone, nell'illusione della 'macchia d'inchiostro' (fig. 6), vedono delle D, e non delle B, se l'illusione è presentata senza preparazione. Anche la funzionalità (un altro aspetto dell'esperienza) influenza le risposte percettive distorte: a paragone con persone che hanno una preparazione musicale bassa o nulla, i musicisti, che hanno necessità di suonare in maniera intonata, mostrano non solo una migliore percezione categoriale degli accordi maggiori e minori, ma anche un'amplificata discriminazione tra accordo intonato (*template*, prototipico), e accordo stonato (non prototipico) (Acker et alii, 1995; Barrett, 1997).

Con questo lavoro si è cercato di fare il punto della situazione su un argomento complesso che rappresenta un 'osso duro' della riflessione fonetica. Ci sono, ovviamente, punti di vista contrastanti con quello qui espresso, in particolare tra coloro che sostengono il primato della codifica fonologica. Le prove empiriche portate a supporto delle varie posizioni sono però largamente inferenziali e derivano spesso da discipline esterne alla ricerca fonetica e fonologica. La sfida per fonetisti e fonologi è quella di fornire prove sperimentali che supportino o confutino direttamente le affermazioni fatte. L'applicazione di paradigmi provenienti da altre discipline (ad esempio la psicolinguistica) può avere un'attrattiva immediata, ma spesso finisce per affermare cose già note. Per assicurare un posto di rispetto tra le scienze linguistiche cognitive alla fonetica e alla fonologia, fonetisti e linguisti dovrebbero impegnarsi a sviluppare una metodologia empirica di maggior rilevanza biologica per dare risposte adeguate alle domande relative allo *status* delle unità linguistiche.

## **BIBLIOGRAFIA**

Acker, B.E., Pastore, R.E. & Hall, M.D. (1995), Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception & Psychophysics*, 57(6), 863-874.

Adams, R.B. & Janata, P. (2002), A comparison of neural circuits underlying auditory and visual object categorization, *NeuroImage*, 16, 361-377.

- Barrett, S.E. (1997), *Prototypes in speech perception*, Unpublished Ph.D. thesis, Univ. of Cambridge.
- Bashford, J.A., Meyers, M.D., Brubaker, B.S. & Warren, R.M. (1988), Illusory continuity of interrupted speech: Speech rate determines durational limits, *Journal of the Acoustical Society of America*, 84(5), 1635-1638.
- Benson, R.R., Richardson, M., Whalen, D.H. & Lai, S. (2006), Phonetic processing in areas revealed by sinewave speech and acoustically similar non-speech, *NeuroImage*, 31, 342-353.
- Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A. & Ward, B.D. (2004), Neural correlates of sensory decision processes in auditory object identification, *Nature Neuroscience*, 7(3), 295-301.
- Bregman, A.S. (1981), Asking the ‘‘what for’’ question in auditory perception, in *Perceptual organization* (M. Kubovy & J.R. Pomerantz, editors), Hillsdale, NJ: Lawrence Erlbaum.
- Bregman, A.S. (1990), *Auditory scene analysis: The perceptual organization of sound*, Cambridge, MA: MIT Press.
- Campbell, R., MacSweeney, M. & Waters, D. (2008), Sign language and the brain, *Journal of Deaf Studies and Deaf Education*, 13, 3-20.
- Carlyon, R.P., Deeks, J.M., Norris, D.G. & Butterfield, S. (2002), The continuity illusion and vowel identification, *Acta Acustica united with Acustica*, 88, 408-415.
- Carlyon, R.P., Deeks, J.M., Shtyrov, Y., Grahn, J., Gockel, H., Hauk, O. & Pulvermüller, F. (in press), Changes in the perceived duration of a narrowband sound induced by a preceding stimulus, *Journal of Experimental Psychology: Human Perception and Performance*.
- Carpenter, G.A. & Grossberg, S. (2003). Adaptive resonance theory, in *The handbook of brain theory and neural networks* (M.A. Arbib, editor). Cambridge, MA: MIT Press, 87-90.
- Chomsky, N. & Halle, M. (1968), *The sound pattern of English*, New York: Harper and Row.
- Ciocca, V. & Bregman, A.S. (1987), Perceived continuity of gliding and steady-state tones through interrupting noise, *Perception & Psycho-physics*, 42, 476-484.
- Cooke, M.P. & Ellis, D.P.W. (2001), The auditory organization of speech and other sources in listeners computational models, *Speech Communication*, 35, 141-177.
- Cooke, M.P. (2006), A glimpsing model of speech perception in noise, *Journal of the Acoustical Society of America*, 119, 1562-1573.
- Denes, P.B. (1963), On the statistics of spoken English, *Journal of the Acoustical Society of America*, 35(6), 892-904.
- Deutsch, D. (editor) (1999), *The psychology of music (2nd ed)*, San Diego: Academic Press.
- Deutsch, D. (2007), Music perception, *Frontiers of Bioscience*, 12, 4473-4482.

- Diehl, R., Lotto, A.J. & Holt, L.L. (2004), Speech perception, *Annual Review of Psychology*, 55, 149-179.
- Elman, J.L. (2004), An alternative view of the mental lexicon, *Trends in Cognitive Sciences*, 8(7), 301-306.
- Fishbach, A., Nelken, I. & Yeshurun, Y. (2001), Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transient, *Journal of Neurophysiology*, 85, 2303-2323.
- Fixmer, E. & Hawkins, S. (1998), The influence of quality of information on the McGurk effect, *Paper presented at the international conference on auditory-visual speech processing*, Terrigal, Australia.
- Fougeron, C. & Keating, P.A. (1997), Articulatory strengthening at edges of prosodic domains, *Journal of Acoustical Society of America*, 101(6), 3728-3740.
- Fowler, C.A. (1986), An event approach to the study of speech-perception from a direct realist perspective, *Journal of Phonetics*, 14(1), 3-28.
- Fowler, C.A. & Smith, M.R. (1986), Speech perception as 'vector analysis': An approach to the problems and segmentation, in *Invariance and variability in speech processes* (J.S. Perkell & D.H. Klatt, editors), Hillsdale: Lawrence Erlbaum Associates, 123-139.
- Fritz, J., Elhilali, M. & Shamma, S. (2005), Active listening: Task-dependent plasticity of spectrotemporal fields in primary auditory cortex, *Hearing Research*, 206, 159-176.
- Fritz, J., Shamma, S., Elhilali, M. & Klein, D. (2003), Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex, *Nature Neuroscience*, 6(11), 1216-1223.
- Gagné, J.-P., Rochette, A.-J. & Charest, M. (2002), Auditory, visual and audiovisual clear speech, *Speech Communication*, 37, 213-230.
- Gaskell, M.G. & Marslen-Wilson, W. (2002), Representation and competition in the perception of spoken words, *Cognitive Psychology*, 45, 220-266.
- Ghazanafar, A.A. & Schroeder, C.E. (2006), Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10, 278-285.
- Green, P.K., Kuhl, K.P. & Meltzoff, N.A. (1988), Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment, *Journal of Acoustical Society of America*, 84, S155.
- Greenberg, S. (2006), A multi-tier framework for understanding spoken language, in *Listening to speech: An auditory perspective* (S. Greenberg & W. Ainsworth, editors), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gregory, R.L. (2005), The Medawar Lecture: Knowledge for vision: Vision for knowledge, *Philosophical Transactions of the Royal Society*, 360 (1458), 1231-1251.
- Griffiths, T.D. & Warren, J.D. (2002), The planum temporale as a computational hub, *Trends in Neurosciences*, 25, 348-353.
- Griffiths, T.D. & Warren, J.D. (2004), What is an auditory object? *Nature Reviews Neuroscience*, 5, 885-890.

- Grossberg, S. (2003), Resonant neural dynamics of speech perception, *Journal of Phonetics*, 31, 423-445.
- Grossberg, S. (2005), Linking attention to learning, expectation, competition, and consciousness, in *Neurobiology of attention* (L. Itti, G. Rees & J. Tsotsos, editors), San Diego: Elsevier, 652-662.
- Halle, M. & Stevens, K.N. (1962), Speech recognition: A model and a program for research, *IRE Transactions on Information Theory*, IT-8(2), 155-159.
- Hawkins, H.L., Hillyard, S.A., Luck, S.J., Mouloua, M., Downing, C.J. & Woodward, D. P. (1990), Visual attention modulates signal detectability, *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 802-811.
- Hawkins, J. (2004), *On intelligence*, New York: Henry Holt and Company (An Owl Book).
- Hawkins, S. (2003), Roles and representations of systematic fine phonetic detail in speech understanding, *Journal of Phonetics*, 31, 373-405.
- Hawkins, S. (2004), Puzzles and patterns in 50 years of research on speech perception, in *From sound to sense: 50+ years of discoveries in speech communication* (J. Slifka, S. Manuel, J. Perkell & S. Shattuck-Hufnagel, editors), Cambridge, MA: MIT.
- Hawkins, S. (in press), Phonetic variation as communicative system: Perception of the particular and the abstract, in *Papers in Laboratory Phonology X* (C. Fougerson, M. D'Imperio, B. Kühnert & N. Vallée, editors), Berlin: Mouton de Gruyter.
- Hawkins, S. & Local, J.K. (2007). Sound to sense: Introduction to the special session, in *16th international congress of phonetic sciences* (W.J. Barry & J. Trouvain, editors). Saarbrücken, 181-184, Paper ID 1726.
- Hawkins, S. & Nguyen, N. (2004), Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English, *Journal of Phonetics*, 32(2), 199-231.
- Hawkins, S. & Smith, R.H. (2001), Polysp: A polysystemic, phonetically-rich approach to speech understanding, *Italian Journal of Linguistics–Rivista di Linguistica*, 13, 99-188.
- Heinrich, A., Carlyon, R.P., Davis, M.H. & Johnsrude, I.S. (2008), Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20(6), 1-16.
- Husain, F.T., Tagamets, T.-A., Fromm, S., Braun, A. & Horwitz, B. (2004), Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study, *NeuroImage*, 21, 1701-1720.
- Iverson, P., Bernstein, L.E. & Auer, E.T. (1998), Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition, *Speech Communication*, 26, 45-63.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S. & Dupoux, E. (2003), Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study, *Journal of Neuroscience*, 23(29), 9541-9546.
- Jakobson, R., Fant, C.G.M. & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*, Cambridge, MA: MIT.

- Jazayeri, M. & Movshon, J. (2007), A new perceptual illusion reveals the mechanisms of sensory decoding, *Nature*, 446(7138), 912-915.
- Kalikow, D.N., Stevens, K.N. & Elliott, L.L. (1977), Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability, *Journal of the Acoustical Society of America*, 61(5), 1337-1361.
- Kanizsa, G. & Gerbino, W. (1982), Amodal completion: Seeing or thinking? In *Organization and representation in perception* (J. Beck, editor), Hillsdale, NJ: Lawrence Erlbaum, 167-190.
- Khalfa, S., Bougeard, R., Morand, N., Veillet, E., Isnard, J., Guenot, M. et al. (2001), Evidence of peripheral auditory activity modulation by the auditory cortex in humans, *Neuroscience*, 104(2), 347-358.
- Kingston, J. & Diehl, R.L. (1994), Phonetic knowledge, *Language*, 70(3), 419-454.
- Klatt, D.H. (1979), Speech perception: A model of acoustic-phonetic analysis and lexical access, *Journal of Phonetics*, 7, 279-312.
- Local, J.K. (2003), Variable domains and variable relevance: Interpreting phonetic exponents, *Journal of Phonetics*, 31, 321-339.
- MacDonald, J. & McGurk, H. (1978), Visual influences on speech perception, *Perception and Psychophysics*, 24(3), 253-257.
- Mattys, S., White, L. & Mehlor, J.F. (2005), Integration of multiple speech segmentation cues: A hierarchical framework, *Journal of Experimental Psychology, General*, 134(4), 477-500.
- McDonough, J. (2001), Incorporating onsets in Navajo: The d-effect, in *Papers in Honour of Ken Hale*, MIT Working Papers in Linguistics (A. Carnie, E. Jelinek & M.A. Willie, editors), 177-188.
- McGurk, H. & MacDonald, J. (1976), Hearing lips and seeing voices, *Nature*, 264, 746-748.
- Meng, M., Remus, D.A. & Tong, F. (2005), Filling-in of visual phantoms in the human brain, *Nature Neuroscience* 8(9), 1248-1254.
- Miller, G.A., Heise, G.A. & Lichten, W. (1951), The intelligibility of speech as a function of the context of the test materials, *Journal of Experimental Psychology*, 41, 329-335.
- Munhall, K.G., Gribble, P., Sacco, L. & Ward, M. (1996), Temporal constraints on the McGurk Effect, *Perception and Psychophysics*, 58(3), 351-362.
- Murray, M.M., Molholm, S., Michel, C., Heslenfeld, D.J., Ritter, W., Javitt, D.C., et al. (2005), Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment, *Cerebral Cortex*, 15, 963-974.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P. & Winkler, I. (2001), 'Primitive intelligence' in the auditory cortex, *Trends in Neurosciences*, 24(5), 283-288.
- Nelken, I., Fishbach, A., Las, L., Ulanovsky, N. & Farkas, D. (2003), Primary auditory cortex of cats: Feature detection or something else? *Biological Cybernetics*, 89(5), 397-406.

- Nguyen, N., Wauquier-Gravelines, S. & Tuller, B. (in press), The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories, in *Approaches to phonological complexity* (I. Chitoran, C. Coupé, E. Marsico & F. Pellegrino, editors), Berlin: Mouton de Gruyter.
- Obleser, J., Scott, S.K. & Eulitz, C. (2006), Now you hear it, now you don't: Transient traces of consonants and their nonspeech analogues in the human brain, *Cerebral Cortex*, 16, 1069-1076.
- Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T. & McCarthy, G. (2005), Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements, *Cerebral Cortex*, 15(12), 1866-1876.
- Pinna, B. & Grossberg, S. (2006), Logic and phenomenology of incompleteness in illusory figures: New cases and hypotheses, *Psychofonia*, IX(15), 93-135.
- Raizada, R.D.S. & Poldrack, R.A. (2007), Selective amplification of stimulus differences during categorical processing of speech, *Neuron*, 56, 726-740.
- Roy, D. (2005a), Grounding words in perception and action: Computational insights, *Trends in Cognitive Sciences*, 9(8), 389-396.
- Roy, D. (2005b), Semiotic schemas: A framework for grounding language in action and perception, *Artificial Intelligence*, 167(1-2), 170-205.
- Sams, M., Manninen, P., Surakka, V., Helin, P. & Kaetoe, R. (1998), McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context, *Speech Communication*, 26(1-2), 75-87.
- Samuel, A. G. (1981), Phonemic restoration: Insights from a new methodology, *Journal of Experimental Psychology: General*, 110, 474-494.
- Samuel, A. G. (1987), Lexical uniqueness effects on phonemic restoration, *Journal of Memory and Language*, 26, 36-56.
- Samuel, A. G. (1997), Lexical activation produces potent phonemic percepts, *Cognitive Psychology*, 32, 97-127.
- Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T. & Munhall, K. (2003), Perceiving biological motion: Dissociating visible speech from walking, *Journal of Cognitive Neuroscience*, 15(6), 800-809.
- Scott, S.K. (2005), Auditory processing – speech, space and auditory objects, *Current Opinion in Neurobiology*, 15, 197-201.
- Seitz, A.R., Nanez, J.E., Holloway, S.R., Koyama, S. & Watanabe, T. (2005), Seeing what is not there shows the costs of perceptual learning, in *Proc. of the National Academy of Science*, 102(25), 9080-9085.
- Sekiyama, K. & Tohkura, Y. (1991), McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility, *Journal of the Acoustical Society of America*, 90(4), 1797-1805.
- Sekiyama, K. & Tohkura, Y. (1993), Inter-language differences in the influence of visual cues in speech perception, *Journal of Phonetics*, 21, 427-444.

- Sekiyama, K. (1996), Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects, *Perception and Psychophysics*, 59, 73.
- Sprague, N., Ballard, D. & Robinson, A. (2007), Modeling embodied visual behaviors, *ACM Transactions on Applied Perception*, 4(2), Article 11.
- Stevens, K.N. (1989), On the quantal nature of speech, *Journal of Phonetics*, 17, 3-45.
- Stevens, K.N. (2002), Toward a model for lexical access based on acoustic landmarks and distinctive features, *Journal of the Acoustical Society of America*, 111, 1872-1891.
- Stevens, K.N. & Halle, M. (1967), Remarks on analysis by synthesis and distinctive features, in *Models for the perception of speech and visual form* (W. Wathen-Dunn, editor), Cambridge, MA: MIT Press, 88-102.
- Stevens, K.N. & Blumstein, S.E. (1978), Invariant cues for place of articulation in stop consonants, *Journal of the Acoustical Society of America*, 64, 1358-1368.
- Studdert-Kennedy, M. & Shankweiler, D. (1970), Hemispheric specialization for speech perception, *Journal of the Acoustical Society of America*, 48, 579-594.
- Sussman, E., Winkler, I., Huottilainen, M., Ritter, W. & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory information, *Cognitive Brain Research*, 13, 393-405.
- Tabor, W., Juliano, C. & Tanenhaus, M.K. (1997), Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing, *Language and Cognitive Processes*, 12(2/3), 211-271.
- Tuller, B. (2003), Computational models in speech perception, *Journal of Phonetics*, 31, 503-507.
- Warren, R.M. (1999), *Auditory perception: A new analysis and synthesis*, Cambridge: Cambridge University Press.
- Weinberger, N.M. (2004), Specific long-term memory traces in primary auditory cortex, *Nature Reviews Neuroscience*, 5, 279-290.
- Werker, J.F. & Tees, R.C. (1984), Phonemic and phonetic factors in adult cross-language speech perception, *Journal of the Acoustical Society of America*, 75, 1866-1878.
- Winer, J.A. (2006), Decoding the auditory corticofugal systems, *Hearing Research*, 212, 1-8.
- Winer, J.A. & Lee, C.C. (2007), The distributed auditory cortex, *Hearing Research*, 229, 3-13.
- Wurm, L.H. (1997), Auditory processing of prefixed English words is both continuous and decompositional, *Journal of Memory and Language*, 37, 438-461.