



# Percezione del linguaggio

▲ Che cosa succede qui? È lecito ritenere che le emozioni espresse dalle facce delle persone siano una risposta a qualcosa che qualcuno abbia detto. La nostra abilità nel percepire i suoni creati dal nostro apparato vocale come parole significative è uno dei nostri più importanti mezzi di comunicazione. Così come gli altri tipi di percezione di cui abbiamo parlato, la facilità di percezione della parola è il risultato finale di processi complessi, che coinvolgono spesso meccanismi cognitivi basati sulle proprie esperienze passate.

**N**onostante percepiamo facilmente il linguaggio

nella maggior parte delle condizioni, sotto questa disinvoltura si annidano processi così complessi come quelli coinvolti nella percezione delle più complicate scene visive. Un modo per apprezzare questa complessità sta nel considerare i tentativi di utilizzare i computer per riconoscere il parlato. Molte compagnie usano ora sistemi di riconoscimento del linguaggio per offrire servizi come la prenotazione di biglietti, servizi bancari automatizzati e supporto tecnico per computer. Ma se hai mai usato uno di questi sistemi, è probabile che un'amichevole voce del computer ti abbia risposto "non capisco che cosa hai detto" in più di un'occasione.

Il riconoscimento vocale del computer è in costante miglioramento, ma ancora non può ancora eguagliare le capacità delle persone di riconoscere il linguaggio. I computer funzionano bene quando una persona parla lentamente e chiaramente e quando non c'è alcun rumore di sottofondo. Invece, gli umani riescono a percepire il parlato in un'ampia varietà di condizioni, inclusa la presenza di diversi rumori di sottofondo, pronuncia disattenta, interlocutori con differenti dialetti e accenti e il botta e risposta spesso caotico che si verifica di routine quando le persone parlano tra di loro (Sinha, 2002; Zue & Glass, 2000). Questo capitolo ti aiuterà ad apprezzare i problemi percettivi complessi posti dal parlato e descriverà la ricerca che ci ha aiutato ad iniziare a capire come il sistema di percezione del linguaggio umano abbia risolto alcuni di questi problemi.

## Lo stimolo del linguaggio

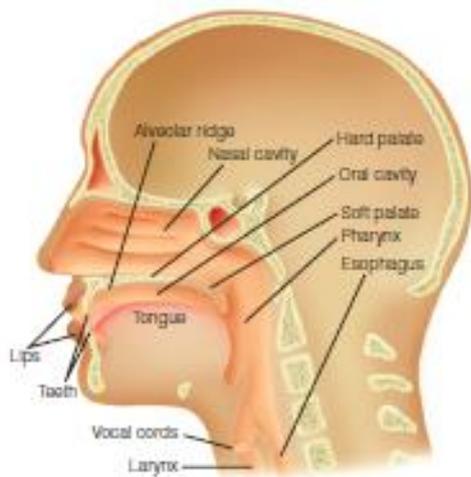
Abbiamo iniziato la descrizione del suono nel capitolo 11 introducendo i toni puri – semplici pattern ad onde sinusoidali con diverse ampiezze e frequenze. Poi abbiamo introdotto i toni complessi consistenti in un numero di toni puri, chiamati armonici, con frequenze che sono multipli della frequenza fondamentale del tono. I suoni del linguaggio aumentano la complessità di un ulteriore livello. Possiamo ancora descrivere il linguaggio in termini di frequenze, ma anche in termini di bruschi inizi e interruzioni, silenzi e rumori che avvengono mentre l'oratore pronuncia le parole. Sono queste parole che aggiungono un'importante dimensione al parlato – i significati che gli interlocutori creano dicendo parole e legandole insieme nelle frasi. Questi significati influenzano la nostra percezione degli stimoli ricevuti, così che quello che percepiamo dipenda non solo dagli stimoli sonori fisici ma anche dai processi cognitivi che ci aiutano ad interpretare quello che sentiamo. Iniziamo descrivendo gli stimoli sonori fisici, chiamati *segnali acustici*.

### I segnali acustici

I suoni del parlato sono prodotti attraverso la posizione o il movimento di strutture interne all'apparato vocale, che produce modelli di variazioni di pressione nell'aria chiamati **stimoli acustici**, o **segnale acustico**. Il segnale acustico per la maggior parte dei suoni del linguaggio è creato dall'aria che viene spinta verso l'alto dai polmoni oltre alle corde vocali nel tratto vocale. Il suono che viene prodotto dipende dalla forma del tratto vocale mentre l'aria viene premuta attraverso di esso. La forma del tratto vocale viene alterata muovendo gli **articolatori**, che includono strutture come la lingua, le labbra, i denti, il mento e il palato molle (**Figura 13.1**).

Consideriamo innanzitutto la produzione delle vocali. Le vocali sono pronunciate tramite la vibrazione delle corde vocali, e il suono specifico di ogni vocale è creato dal cambiamento della forma generale del tratto vocale. Questo cambiamento in forma modifica la frequenza risonante del tratto vocale e produce picchi di pressione a numeri di frequenze diverse (**Figura 13.2**). Le frequenze alle quali si verificano questi picchi sono chiamate **formanti**.

Ogni suono vocalico ha una serie caratteristica di formanti. Il primo formante ha la frequenza più bassa; il secondo formante è il successivo più alto, e così via. I formanti delle vocali /ae/ (il suono vocalico nella parola *had*) sono mostrati su uno **spettrogramma del suono** in **Figura 13.3** (i suoni del parlato sono indicati con

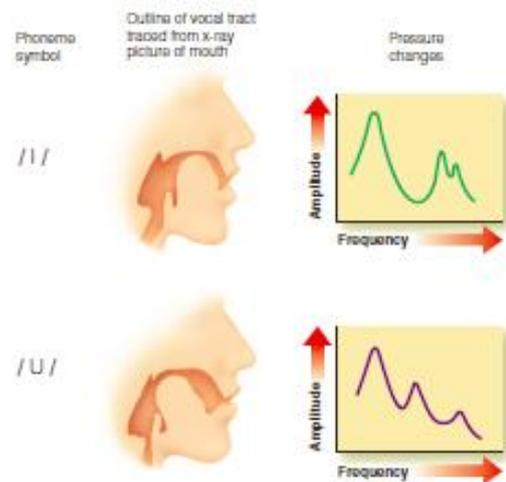


**Figura 13.1** Il tratto vocale include le cavità nasali e orali e la faringe, così come componenti che si muovono, come la lingua, le labbra e le corde vocali.

delle barre). Lo spettrogramma del suono indica il modello di frequenze e intensità nel corso del tempo che costituiscono il segnale acustico. La frequenza è indicata sull'asse verticale e il tempo sull'asse orizzontale; l'intensità è indicata dalla parte nera, di cui le parti più scure indicano una maggiore intensità. Dalla Figura 13.3 possiamo vedere che i formanti sono concentrazioni di energia a frequenze specifiche, con il suono /ae/ avente formanti a 500, 1,700 e 2,500 Hz. Le linee verticali nello spettrogramma sono le oscillazioni di pressione causate dalle vibrazioni delle corde vocali.

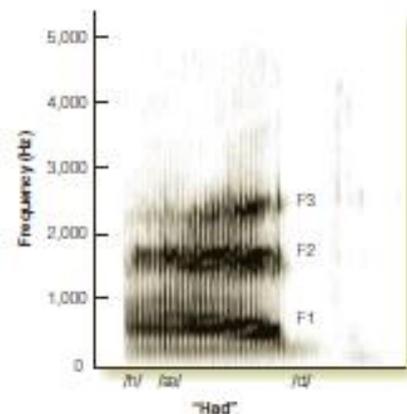
Le consonanti sono prodotte da una costrizione, o chiusura, del tratto vocale. Per illustrare la differenza con cui sono prodotte le consonanti, consideriamo i suoni /d/ e /f/. Fai questi suoni, e nota che cosa la tua lingua, labbra e denti stanno facendo. Quando produci il suono /d/, posizioni la lingua contro la cresta sopra i tuoi denti superiori (la cresta alveolare di Figura 13.1) e poi rilasci un leggero soffio d'aria come muovi la tua lingua via dalla cresta alveolare (provalo). Quando produci il suono /f/, posizioni il labbro inferiore contro i tuoi denti frontali superiori e poi soffi l'aria tra le labbra e i denti.

Questi movimenti della lingua, labbra e altri articolatori creano gli schemi di energia nel segnale acustico che possiamo osservare sullo spettrogramma del suono. Per esempio, lo spettrogramma della frase "Roy and read the will", mostrato in Figura 13.4, mostra gli aspetti del segnale associati con le vocali e consonanti. Le tre bande orizzontali chiamate F1, F2 e F3 sono i tre formanti associati con il suono /e/ di *read*. Rapidi cambiamenti in frequenza che precedono o seguono i formanti sono chiamati **transizioni formanti** e sono associati alle consonanti. Per esempio, T2 e T3 sono transizioni formanti associate con il suono /r/ di *read*.



**Figura 13.2** A sinistra: la forma del tratto vocale per le vocali /i/ (come in *zip*) e /u/ (come in *put*). A destra: l'ampiezza delle variazioni di pressione prodotte per ogni vocale. I picchi nei cambiamenti di pressione sono i *formanti*. Ogni suono vocalico ha un caratteristico pattern di formanti che sono determinati dalla forma del tratto vocale per quella vocale.

Abbiamo descritto le caratteristiche fisiche del *segnale acustico*. Per capire come questo segnale acustico si traduce nella *percezione del linguaggio*, dobbiamo considerare le unità base del parlato.



**Figura 13.3** Spettrogramma della parola *had*. "Tempo" è sull'asse orizzontale. Le bande orizzontali scure sono il primo (F1), il secondo (F2) e il terzo (F3) formanti associati con il suono della vocale /ae/.

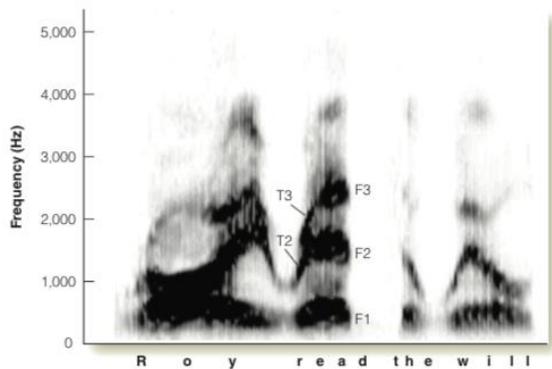
## Unità base del parlato

Il nostro primo compito nello studio della percezione del parlato è quello di separare i suoni del parlato in unità gestibili. Che cosa sono queste unità? Il flusso della frase? Una particolare parola? Una sillaba? Il suono di una lettera? Una frase è un'unità troppo lunga per una semplice analisi, e alcune lettere non hanno alcun suono. Sebbene questi siano argomenti a favore dell'idea che la sillaba sia l'unità base del parlato (Mehler, 1981; Segui, 1984), la gran parte della ricerca nel linguaggio si è basata su un'unità chiamata **fonema**.

Un fonema è il più breve segmento del parlato che, se mutato, cambierebbe il significato di una parola. Considera la parola *bit*, che contiene i fonemi /b/, /i/ e /t/. Sappiamo che /b/, /i/ e /t/ sono fonemi perché possiamo cambiare il significato della parola alterando ogni fonema individualmente. Così, *bit* diventa *pit* se /b/ è cambiato in /p/, *bit* diventa *bat* se /i/ è cambiato in /a/ e *bit* diventa *bid* se /t/ è cambiato in /d/.

I fonemi dell'inglese americano, elencati in **Tabella 13.1**, sono rappresentati da simboli fonici che simboleggiano i suoni del linguaggio. Questa tabella mostra i fonemi per 13 suoni vocalici e 24 suoni consonantici. Ad una prima occhiata alla tabella potresti pensare che ci siano più vocali del set standard che hai imparato alla scuola elementare (*a, e, i, o, u*, e qualche volta *y*). Il motivo per cui ci sono più vocali è che alcune vocali possono avere più di una pronuncia, per cui ci sono più suoni vocalici che lettere vocaliche. Ad esempio, la vocale *o* suona in modo diverso in *boat* e *hot*, e la vocale *e* suona diversamente in *head* e *heed*. I fonemi, quindi, si riferiscono non alle lettere ma ai suoni del linguaggio che determinano il significato di quello che le persone dicono.

Dal momento che diversi linguaggi usano diversi suoni, il numero di fonemi varia tra i linguaggi. Ci sono solo 11 fonemi nell'hawaiano, ma ne sono stati identificati ben 47 nell'inglese americano e fino a 60 in alcuni linguaggi africani. Dunque, i fonemi sono definiti in termini del suono che viene usato per creare le parole in una specifica lingua.



**Fig.13.4** Spettrogramma della frase "Roy legge il di transizione T2 e T3."

CONSONANTS				VOWELS	
p	<i>pull</i>	s	<i>sip</i>	i	<i>heed</i>
b	<i>bull</i>	z	<i>zip</i>	l	<i>hid</i>
m	<i>man</i>	r	<i>rip</i>	e	<i>baft</i>
w	<i>will</i>	ʃ	<i>should</i>	ɛ	<i>head</i>
f	<i>fill</i>	ʒ	<i>pleasure</i>	æ	<i>had</i>
v	<i>vet</i>	ç	<i>chop</i>	u	<i>who'd</i>
θ	<i>thigh</i>	ǰ	<i>gyp</i>	U	<i>put</i>
ð	<i>that</i>	y	<i>yip</i>	ʌ	<i>but</i>
t	<i>tie</i>	k	<i>kale</i>	o	<i>boat</i>
d	<i>die</i>	g	<i>gale</i>	O	<i>bought</i>
n	<i>near</i>	h	<i>hail</i>	a	<i>hot</i>
l	<i>lear</i>	ŋ	<i>sing</i>	ə	<i>sofa</i>
				ɪ	<i>many</i>

There are other American English phonemes in addition to those shown here, and specific symbols may vary depending on the source.

© Cengage Learning

**Tab.13.1** Principali consonanti e vocali dell'inglese e i loro simboli fonetici.

Ci sono altri fonemi dell'inglese americano in aggiunta a quelli mostrati qua e simboli specifici possono variare a seconda della fonte.

Può sembrare che avendo identificato il fonema come unità base del discorso, sia possibile descrivere tale percezione in termine di stringhe di fonemi.

In base a questa idea, percepiamo una serie di suoni chiamati fonemi, che creano sillabe che si combinano in parole.

Queste sillabe e parole ci appaiono unite una dopo l'altra come perle su un filo.

Per esempio, percepiamo la frase "la percezione è semplice" come la sequenza di unità "la-per-ce-zio-ne-è-sem-pli-ce". Ciò nonostante percepire il discorso può sembrare soltanto un problema di processare una serie di suoni separati che sono allineati uno dopo l'altro, in realtà la situazione è molto più complessa.

Piuttosto che susseguirsi, con la fine del segnale di un suono e poi l'inizio del segnale successivo, come lettere su una pagina, segnali per suoni adiacenti si sovrappongono.

Inoltre, l'andamento della pressione dell'aria che cambia per un particolare suono può variare molto se

chi parla è maschio o femmina, giovane o anziano, se parla rapidamente o lentamente o ha un determinato accentto.

## La relazione variabile tra fonema e segnale acustico

Il problema principale che hanno incontrato i ricercatori tentando di capire la percezione del discorso è la presenza di una relazione mutevole tra il segnale acustico e i suoni che sentiamo.

In altre parole, un particolare suono può essere associato a un diverso numero di segnali acustici.

Andiamo a considerare alcune delle cause di questa variabilità.

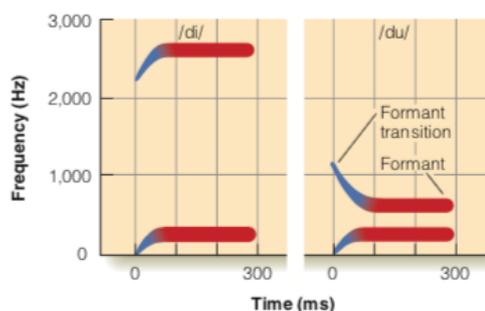
## Variabilità data dal contesto

Il segnale acustico associato a un fonema cambia a seconda del suo contesto.

Per esempio, guarda la Fig. 13.5, che mostra gli spettrogrammi dei suoni /di/ e /du/.

Questi sono spettrogrammi piatti disegnati a mano che mostrano le due più importanti caratteristiche dei suoni: i formanti (in rosso) e i formanti di transizione (in blu).

Dal momento che i formanti sono associati alle vocali, sappiamo che questi a 200 e 2600 Hz sono i segnali acustici della vocale /i/ nel suono /di/ e che i formanti a 200 e 600 Hz sono il segnale acustico della vocale /u/ nel suono /du/.



**Fig. 13.5** Spettrogrammi disegnati a mano per /di/ e /du/.

Poiché i formanti sono i segnali acustici delle vocali, i formanti di transizione che precedono i formanti devono essere i segnali per la consonante /d/.

Tuttavia notiamo che i formanti di transizione per i secondi (ad alta frequenza) per /di/ e /du/, sono diversi. Per /di/, il formante di transizione inizia a circa 2200 Hz e si alza a circa 2600 Hz.

Per /du/, il formante di transizione inizia circa a 1100 Hz e si abbassa a circa 600 Hz.

Pertanto, anche se percepiamo lo stesso suono /di/ in /di/ e /du/, i formanti di transizione, che sono segnali acustici associati a questi suoni sono diversi.

Quindi, il contesto in cui uno specifico fonema si presenta può influenzare il segnale acustico associato a esso.

Questo effetto del contesto si verifica per il modo in cui il discorso viene prodotto. Dal momento che le articolazioni sono in costante movimento mentre parliamo, la forma del tratto della vocale associata ad un particolare fonema è influenzata dai suoni che precedono e seguono questo fonema.

Questa sovrapposizione tra l'articolazione di fonemi adiacenti è chiamata co-articolazione.

Puoi dimostrare la co-articolazione tu stesso notando come produci fonemi in differenti contesti.

Per esempio, pronuncia pipistrello (bat) e stivale (boot). Quando dici pipistrello, le tue labbra non sono arrotondate, invece pronunciando stivale lo sono, anche durante il suono iniziale /b/.

Tuttavia, anche se la /b/ è la stessa in entrambe le parole la articoli in modo diverso. In questo esempio l'articolazione di /oo/ in stivale (boot) si sovrappone all'articolazione di /b/, facendo in modo che le labbra risultino arrotondate anche prima che il suono /oo/ venga prodotto.

Il fatto che percepiamo il suono di un fonema sempre uguale anche se il segnale acustico è cambiato per co-articolazione è un esempio di costanza percettiva. Questo termine potrebbe esserti familiare in base alle nostre osservazioni sul fenomeno della costanza visiva, come la costanza del colore (percepiamo il colore cromatico di un oggetto, come costante anche quando la distribuzione della lunghezza d'onda dell'illuminazione cambia) e

la costanza di grandezza (percepiamo la dimensione di un oggetto come costante anche quando la sua immagine cambia sulla nostra retina). La costanza percettiva è simile come significato. Percepiamo il suono di un particolare fonema anche quando appare in diversi contesti che cambiano il suo segnale acustico.

## Variabilità data da diversi relatori

Le persone pronunciano le stesse parole in diversi modi. Alcuni hanno la voce alta, altri bassa e parlano con vari accenti; alcuni parlano rapidamente e altri e-s-t-r-e-m-a-m-e-n-t-e l-e-n-t-a-m-e-n-t-e. Queste ampie variazioni nei discorsi significano che per diversi relatori, un particolare fonema o parola può avere segnali acustici molto diversi.

Chi parla, inoltre, presenta variabilità attraverso una pronuncia superficiale. Per esempio, pronuncia la seguente frase alla velocità con cui parli ad un amico: "Questo è stato l'acquisto migliore (best buy)". Come hai pronunciato "acquisto migliore (best buy)"? Hai pronunciato la /t/ di migliore (best), o hai detto "best buy"? Che ne pensi di "She is a bad girl (è una cattiva ragazza)"? Dicendolo rapidamente, nota se la sua lingua tocca il palato dicendo la /d/ di "bad".

Infine, cosa ne pensi di "Did you go to the store (sei andato al negozio)?" Hai detto "did you" o "dijoo"? Hai i tuoi modi di pronunciare varie parole e fenomeni e altre persone hanno i loro. Studi su come le persone parlano, in realtà hanno stabilito che ci sono 50 modi diversi di produrre il suono "the" (Waldrop, 1988).

Quelle persone di solito non articolano ciascuna parola individualmente nel discorso colloquiale e questo viene rispecchiato nello spettrogramma in Fig. 13.6.

Lo spettrogramma rappresentato in Fig. 13.6a si riferisce alla domanda "What are you doing? (Che cosa stai facendo?)" pronunciata lentamente e chiaramente.

Lo spettrogramma nella Fig. 13.6b si riferisce alla stessa domanda posta in linguaggio non-formale nel quale "What are you doing?" diventa "Whad'aya doin?"

Questa differenza è mostrata chiaramente nello spettrogramma.

Nonostante le prime e le ultime parole (what e doing) creino dei pattern simili nei due spettrogrammi le pause tra parole sono assenti o molto ovvie nello spettrogramma della Fig. 13.6b, il mezzo di questo spettrogramma è completamente cambiato con un numero di suoni dei discorsi che mancano.

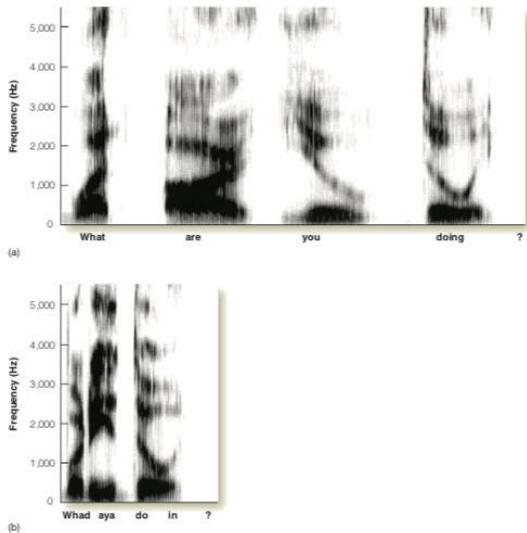
La variabilità nel segnale acustico causata dalla co-articolazione, da differenti speakers e da una pronuncia scorretta crea un problema per l'ascoltatore che deve in qualche modo trasformare l'informazione contenuta in questo segnale acustico con alta variabilità in parole familiari.

Nella prossima sezione considereremo alcuni dei modi dei sistemi di percezione del linguaggio in accordo con il problema della variabilità.

## Fonemi percettivi

Il sistema di percezione del linguaggio è in accordo con il problema di variabilità in diversi modi.

Noi prima descriveremo una proprietà del sistema del linguaggio chiamata percezione categoriale e poi considereremo come l'informazione ricavata dal volto e dalla nostra conoscenza del linguaggio ci aiuta a percepire accuratamente i suoni del discorso.



**Fig. 13.6**  
**(a)** Spettrogramma di “What are you doing?” pronunciato piano e chiaramente.  
**(b)** Spettrogramma di “What are you doing?” pronunciato nel linguaggio colloquiale.

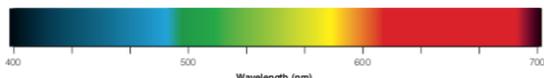
## Percezione categoriale

La percezione categoriale si presenta quando gli stimoli esistenti lungo un continuum sono percepiti come divisi in categorie distinte.

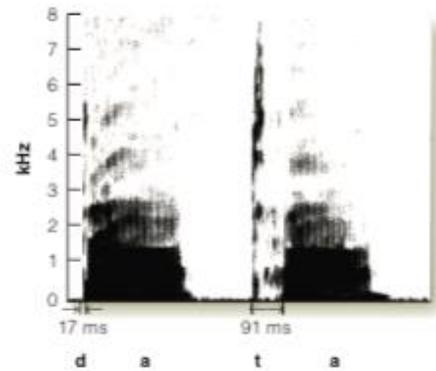
Per esempio, considerando lo spettro visibile in Fig. 13.7. Iniziando da sinistra, a una lunghezza d’onda di 450nm vediamo blu. Mentre se ci muoviamo su lunghezze d’onda più ampie, il colore rimane blu, finché improvvisamente intorno ai 480nm percepiamo verde. Muovendoci lungo il continuum, vediamo verde fino a 750nm, dove il colore cambia e diventa prima giallo, poi arancione e infine rosso.

Così muovendoci lungo l’intero tratto dello spettro visibile ci imbattiamo solo in cinque categorie.

La percezione categoriale nel linguaggio si presenta nello stesso modo, solamente che il continuum è sostituito da una proprietà detta “Attacco della sonorità” “voice onset time (VOT)”; il tempo trascorso da quando un suono inizia e quando le corde vocali iniziano a vibrare. Possiamo illustrare questo intervallo comparando gli spettrogrammi dei suoni /da/ e /ta/ in Fig. 13.8 .



**Fig. 13.7** La percezione categoriale dei colori nello spettro visibile. Lungo il range della lunghezza d’onda sono percepite cinque categorie di colori: blu, verde, giallo, arancione e rosso.



**Fig. 13.8** Spettrogrammi per /da/ e /ta/. Il tempo d’inizio della voce –il tempo trascorso tra l’inizio del suono e l’inizio della sonorizzazione- è indicato all’inizio dello spettrogramma per ogni suono.

Questi spettrogrammi mostrano che il tempo trascorso tra l’inizio del suono e l’inizio della vibrazione delle corde vocali (indicato dalla presenza di linee verticali nello spettrogramma) è di 17ms per /da/ e 91ms per /ta/. Pertanto /da/ ha una corta (VOT) e /ta/ ha una lunga (VOT).

Usando i computer, i ricercatori hanno creato stimoli sonori in cui il VOT viene variato in piccoli step da corto a lungo. Quando variamo il VOT, usando stimoli come quelli mostrati nella Fig. 13.8 e chiediamo agli ascoltatori di indicare quale suono sentono, gli ascoltatori riportano di sentire soltanto uno o l’altro dei due fonemi, /da/ o /ta/, anche se vengono presentati un largo numero di stimoli con VOT differenti.

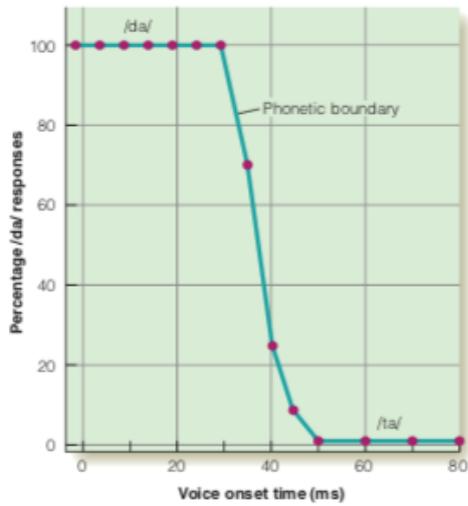
Questo risultato è mostrato nella Fig. 13.9 (Eimas e Corbit, 1973). Con VOT corti, gli ascoltatori riportano di sentire /da/ e continuano a riportare ciò anche quando il VOT viene alzato.

Quando il VOT raggiunge circa 35ms, la loro percezione cambia all’improvviso, così a VOT oltre 40ms, gli ascoltatori riferiscono di sentire /ta/. Il VOT quando una percezione cambia da /da/ a /ta/ è chiamato limite fonetico. Il risultato chiave dell’esperimento sulla percezione categoriale è che anche se il VOT viene variato continuamente attraverso un largo intervallo, l’ascoltatore percepisce solo 2 categorie: /da/ da un lato del limite fonetico e /ta/ dall’altra parte.

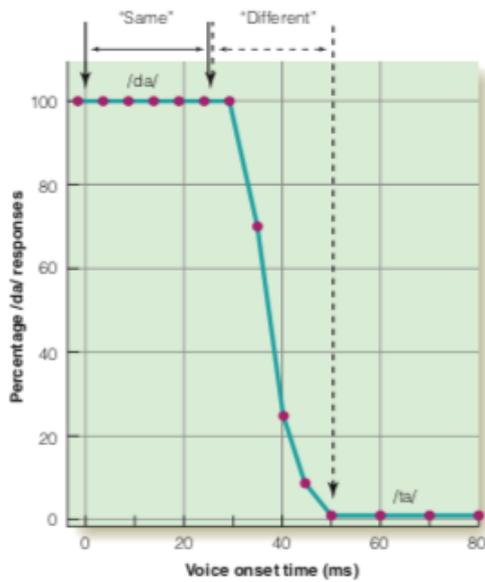
Una volta che abbiamo dimostrato la percezione categoriale usando questa procedura, possiamo passare a un test di discriminazione, in cui presentiamo due stimoli con differenti VOT e chiediamo all’ascoltatore se il suono è lo stesso o diverso. Quando presentiamo due stimoli separati da un VOT di 25ms che sono dalla stessa parte del limite fonetico, come stimoli di 0ms e 25ms, l’ascoltatore afferma che sembrano uguali (Fig. 13.10).

Pertanto, quando presentiamo due stimoli che sono separati dalla stessa distanza in VOT, ma sono su lati opposti del limite fonetico, come stimoli con VOT di 25ms e 50ms, l’ascoltatore dice che sono diversi. Il fatto che tutti gli stimoli sullo stesso lato del limite fonetico siano percepiti come la stessa categoria è un esempio di costanza percettiva. Se questa costanza non esistesse, percepiremmo suoni differenti ogni volta che cambiano il VOT. Invece, sentiamo un suono su ogni

lato del limite fonetico. Questo semplifica la nostra percezione dei fonemi e ci aiuta a percepire più facilmente la vasta varietà di suoni nel nostro ambiente.



**Fig. 13.9** I risultati dell'esperimento sulla percezione categoriale indicano un limite fonetico, con /da/ percepito per VOT a sinistra e /ta/ percepito per VOT a destra.



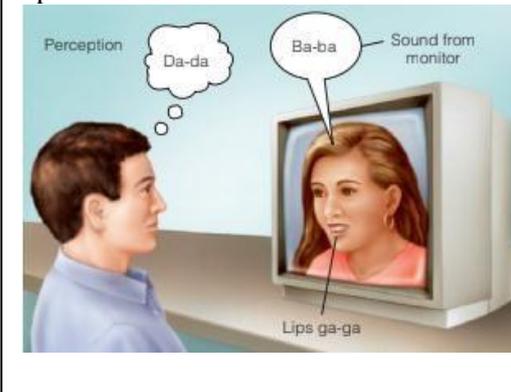
**Fig. 13.10** Nella parte di discriminazione di un esperimento di percezione categoriale, vengono presentati più stimoli e l'ascoltatore indica se sono uguali o diversi. Il tipico risultato è che due stimoli con VOT sullo stesso lato del limite fonetico (VOT=0ms e 25ms; frecce continue) sono giudicati come lo stesso, invece due stimoli su diversi lati del limite fonetico (VOT=25ms e 50ms; frecce tratteggiate) sono giudicati come diversi.

Se questa costanza non esistesse, percepiremmo suoni diversi ogni volta che modifichiamo il VOT. Invece, facciamo esperienza di un suono ad ogni lato del confine fonetico. Ciò semplifica la nostra percezione dei fonemi e ci aiuta a percepire più facilmente l'ampia varietà di suoni presenti nel nostro ambiente.

## Informazioni fornite dal volto

Un'altra proprietà della percezione del parlato è il suo essere multimodale; ciò significa che la nostra percezione del discorso può essere influenzata da informazioni provenienti da numerosi sensi. Un esempio di come la percezione del discorso possa essere influenzata da informazioni visive è mostrato nella figura 13.11. Inizialmente l'ascoltatore sente i suoni /ba-ba/ provenienti dal monitor. Quando è aggiunta la stimolazione visiva, sottoforma di video che mostra una persona i cui movimenti labiali sono quelli relativi al suono /ga-ga/, l'ascoltatore inizia a sentire il suono /da-da/. Nonostante il partecipante stia ancora ricevendo il segnale acustico /baba/, la sua percezione è mutata cosicché egli senta /da-da/. Questo effetto è chiamato effetto McGurk, dal nome del primo autore a descriverlo (McGurk & MacDonald, 1976). Egli dimostra che nonostante l'informazione uditiva sia la maggiore fonte di informazione per la percezione del parlato, anche l'informazione visiva può esercitare una forte influenza su ciò che sentiamo. Questa influenza che la vista ha sulla percezione del discorso è chiamata percezione audiovisiva del parlato. L'effetto McGurk è un esempio di percezione audiovisiva del parlato. Un altro esempio è il modo in cui le persone usano quotidianamente informazioni fornite dal labiale dell'interlocutore per aiutarsi a comprendere il discorso in un contesto rumoroso (vedi anche Sumbly & Pollack, 1954).

**Fig. 13.11. Effetto McGurk.** Le labbra della donna si muovono dicendo ga-ga. Il suono proveniente dagli altoparlanti del monitor è Ba-ba. La persona che vede e ascolta percepisce Da-da. Se la persona chiude gli occhi in maniera da non vedere più la bocca muoversi percepisce ga-ga. Quindi ciò che l'ascoltatore sente dipende dall'informazione visiva.



È stato dimostrato che il legame tra vista e discorso ha una base fisiologica. Gemma Calvert e collaboratori (1997) utilizzò la Risonanza Magnetica Funzionale (fMRI) per misurare l'attività cerebrale mentre i partecipanti guardavano un video silenzioso di una persona intenta a riprodurre i movimenti labiali necessari per dire i numeri. Gli osservatori ripetevano silenziosamente i numeri mentre guardavano, quindi questo compito è simile a ciò che fanno le persone quando leggono il labiale. In una condizione di controllo, gli osservatori guardavano un volto fermo mentre ripetevano i numeri. Il confronto dell'attività cerebrale in queste due condizioni mostrò che guardare il movimento delle labbra attiva un'area nella corteccia uditiva che è attiva quando gli individui percepiscono un discorso, come dimostrato da Calvert in un altro esperimento.

Il fatto che le stesse aree si attivano per la lettura del labiale e la percezione del discorso, ipotizza Calvert, potrebbe essere il meccanismo neurale alla base dell'effetto McGurk.

Il legame tra percezione del parlato e dei volti è stato dimostrato in un altro modo da Katharina von Kriegstein e collaboratori (2005), misurando la Risonanza Magnetica Funzionale (fMRI) mentre gli ascoltatori eseguivano una serie di compiti che comprendevano frasi pronunciate da interlocutori conosciuti (altre persone che lavoravano nel laboratorio) e sconosciuti (persone che non avevano mai sentito parlare prima). Semplicemente ascoltare il discorso attiva il solco temporale superiore (STS; vedi Figura 5.52), un'area che è stata associata alla percezione del parlato in studi precedenti (Belin et al., 2000). Ma quando ai partecipanti viene chiesto di completare compiti che

implicano di prestare attenzione al suono di voci familiari, si attiva anche l'area fusiforme facciale (FFA). All'opposto, prestare attenzione al suono di voci sconosciute non attiva tale area. Apparentemente, quando le persone sentono una voce che associano ad una specifica persona, questo attiva aree non dedicate unicamente alla percezione del parlato, ma anche alla percezione dei volti.

Il collegamento tra la percezione del discorso e quella dei volti, dimostrato sia in esperimenti comportamentali che fisiologici, fornisce informazioni che ci aiutano a gestire la variabilità dei fonemi (vedi Hall et al., 2005, e Wassenhove et al., 2005, sul legame tra osservare qualcuno parlare e percepire il discorso.)

## Informazioni provenienti dalla nostra coscienza del linguaggio

Numerose ricerche dimostrano che è più facile percepire i fonemi ricevuti in un contesto dotato di significato. Philip Rubin, M. T. Turvey, e Peter Van Gelder (1976), ad esempio, presentarono una serie di parole brevi, come "sin", "bat", e "leg", o non-parole, come "jum", "baf", e "teg", e chiesero ai partecipanti di rispondere premendo un tasto il più velocemente possibile quando sentivano un suono cominciava con /b/. In media, i partecipanti impiegavano 631 ms per reagire alle nonparole e 580 ms per rispondere alle parole vere. Quindi, quando il fonema è all'inizio della parola vera, viene identificato 8 volte più velocemente di quando è presentato all'inizio di una sillaba priva di significato.

L'effetto del significato sulla percezione dei fonemi è stato dimostrato ulteriormente da Richard Warren (1970), che chiese ai partecipanti di ascoltare la seguente frase registrata: "I governatori si sono incontrati, con le rispettive legislature, nella capitale." Warren sostituì la prima /s/ in "legislature" con un colpo di tosse e chiese ai partecipanti di indicare dove, nella frase, ci fosse stato il suddetto suono. Nessuno dei partecipanti identificò la corretta posizione, ancor più significativamente nessuno notò che la /s/ in "legislature" mancasse. Di questo effetto, che Warren chiamò l'effetto di restauro fonemico, fecero esperienza anche gli studenti e lo staff del dipartimento di psicologia i quali erano a conoscenza dell'assenza della /s/.

Warren non solo dimostrò l'effetto di restauro fonemico, ma provò anche che può essere influenzato dal significato delle parole successive al fonema mancante. Per esempio, l'ultima parola della frase "There was time to \*ave . . ." (dove \* indica la presenza di un colpo di tosse o un altro suono) potrebbe essere "shave," "save," "wave," o "rave," ma i partecipanti sentono la parola "wave" ("salutare") quando il contesto della frase riguardava un amico in partenza.

L'effetto di restauro fonemico venne usato da Arthur Samuel (1981) per dimostrare che la percezione del parlato è determinata sia dalla natura del segnale acustico (processamento bottom-up) che dal contesto che produce aspettative nell'ascoltatore (processamento top-down). Samuel dimostrò il processamento bottom-up provando che il restauro è migliore quando un suono mascherante, come il ronzio di una televisione accesa su un canale non trasmettente, e il fonema coperto hanno un suono simile. Quindi, il restauro fonemico accade più facilmente per un fonema come /s/, ad altra frequenza, se anche la maschera contiene in larga parte suoni ad alta frequenza.

Ciò che accade nel restauro fonemico, secondo Samuel, è che prima di percepire effettivamente un suono "restaurato", la sua presenza deve essere confermata dalla presenza di un suono ad esso simile. Se la maschera di rumore bianco contiene frequenze simili a quelle del fonema che ci aspettiamo, avviene il restauro, ed è probabile sentire il fonema. Se la maschera non è un suono simile, è meno probabile che accada il fenomeno di restauro fonemico (Samuel, 1990).

Samuel dimostrò il processamento top-down provando che le parole più lunghe aumentano la probabilità dell'effetto di restauro fonemico. Apparentemente, i partecipanti utilizzano il contesto aggiuntivo fornito dalla parola lunga per identificare il fonema coperto. Ulteriore prova dell'importanza del contesto è la dimostrazione di Samuel che vi è maggiore restauro per una parola come "prOgresso" (dove la lettera maiuscola indica il fonema coperto) che per parole inventate come "crOgresso" (Samuel, 1990; vedi anche Samuel, 1997, 2001, per ulteriori prove che il processamento top-down è coinvolto nel restauro fonemico).

## Percepire le parole

Così come la percezione dei fonemi va oltre la semplice elaborazione del segnale acustico, la percezione delle parole dipende da numerosi fattori. Inizieremo dimostrando come l'appartenenza di una parola ad una proposizione può influenzare la nostra percezione della stessa. Considereremo in seguito la capacità di distinguere le parole di una frase, e come possiamo percepire le parole anche quando sono pronunciate in modo diverso da interlocutori differenti.

## Percepire le parole in una frase

Un modo di illustrare come l'appartenere ad una proposizione può influenzare la nostra percezione delle parole è mostrando come le parole possono essere lette anche quando incomplete, come nel seguente esempio.

<b>DIMOSTRAZIONE</b>
<b>Percepire Frasi Incomplete</b>
Leggi le seguenti frasi:
1. M*RI*A H* U* PI**O*O AG*E**O I* C*I P**O E' B*A*C*
CO*E *A N*V*
2. I* S*L* O**I NO* S*LE**E
3. A**UN* PA**L* S*N* PI* FA**L* D* CO*P*END**E *I A**RE

---

### DEMONSTRATION

#### Perceiving Degraded Sentences

Read the following sentences:

1. M\*R\* H\*D \* L\*TTL\* L\*MB I'S FL\*\*C\* W'S WH\*T\* \*S SN\*W
  2. TH\* S\*N \*S N\*T SH\*N\*NG T\*D\*\*
  3. S\*M\* W\*\*DS \*R\* EA\*I\*R T\* U\*D\*R\*T\*N\* T\*A\* \*T'E\*S
- 

L'abilità di leggere le frasi, nonostante metà delle lettere siano state eliminate, è coadiuvata dalla conoscenza della lingua italiana, come le parole sono collegate nel formare frasi, e forse nel primo esempio la tua conoscenza della filastrocca, (Denes & Pinson, 1993).<sup>1</sup>

Un simile effetto di significato si presenta anche nel parlato. Una prima dimostrazione di come l'avere significato rende più facilmente percepibili parole pronunciate, è fornita da George Miller e Steven Isard (1963), che dimostrarono che le parole sono più intelligibili quando ascoltate nel contesto di una frase rispetto a quando sono presentate in una lista di parole non correlate. Gli autori dimostrarono ciò presentando tre tipi di stimoli: (1) normali frasi grammaticali, come "I gadgets semplificano i lavori di casa"; (2) frasi anomale che seguono le regole grammaticali ma non hanno senso, come "I gadgets letteralmente, invece che trovare la frase di una filastrocca italiana ugualmente conosciuta.

---

<sup>1</sup> (NdT): la frase nel testo originale è "MARY HAD A LITTLE LAMB

ITS FLEECE IS WHITE AS SNOW", ho preferito tradurlo

uccidono i viaggiatori dagli occhi”; e (3) stringhe di parole sgrammaticate, come per esempio “Tra i gadgets autostrade viaggiatori il furto”.

Miller e Isard hanno utilizzato una tecnica chiamata **shadowing** (= seguire passo passo), nella quale hanno fatto ascoltare queste frasi ai soggetti attraverso degli auricolari e chiesto loro di ripetere ad alta voce ciò che stavano sentendo. I partecipanti hanno ripetuto frasi normali con un’accuratezza pari all’89 per cento, ma la loro accuratezza scese al 79 per cento con frasi anomale e al 56 per cento con stringhe sgrammaticate. Le differenze tra queste tre tipologie di stimoli sono diventate ancora più significative quando gli ascoltatori sentivano gli stimoli stessi in presenza di un rumore di sottofondo. Ad esempio, con un livello di rumore di sottofondo moderatamente elevato, l’accuratezza era 63 per cento per frasi normali, 22 per cento per frasi anomale, e soltanto 3 per cento per stringhe di parole sgrammaticate. Questi risultati ci dicono che quando le parole sono disposte secondo uno schema significativo, siamo in grado di percepirle più facilmente. Ma la maggior parte delle persone non si rende conto del fatto che è proprio la loro conoscenza della natura del loro linguaggio che li aiuta ad integrare suoni e parole che potrebbero essere difficili da sentire. Ad esempio, la nostra conoscenza di strutture di parole ammissibili ci dice che ANT, TAN, e NAT sono tutte sequenze di lettere ammissibili in lingua inglese, ma che TQN o NQT non possono essere parole della lingua inglese.

Un simile effetto del significato sulla percezione si manifesta anche perché la nostra conoscenza delle regole grammaticali ci dice che “There is no time to question” è una frase ammissibile in lingua inglese mentre “Question, no time there is” non è ammissibile o, nel migliore dei casi, è particolarmente strana (a meno che tu non sia Yoda, che dice questo in *Star Wars, Episodio III: La vendetta dei Sith*). Poiché la maggior parte delle volte ci troviamo di fronte a parole di significato compiuto e frasi grammaticalmente corrette, utilizziamo di continuo la nostra conoscenza di ciò che è ammissibile nella nostra lingua per aiutarci a comprendere che cosa viene detto. Questo diventa particolarmente importante quando ascoltiamo in condizioni non ideali, come ambienti rumorosi o quando la qualità della voce o l’accento di colui che parla è difficile da capire (vedi anche Salasoo & Pisoni, 1985).

## Percepire le pause in una sequenza di parole

Così come vediamo degli oggetti senza fare alcuno sforzo quando guardiamo una scena visiva, di solito abbiamo poca difficoltà nel percepire le singole parole mentre conversiamo con un’altra persona. Ma quando osserviamo il segnale vocale, notiamo che il segnale acustico è continuo, senza pause fisiche nel segnale né

pause che corrispondano necessariamente a quelle che percepiamo tra le parole (**Figura 13.12**). La percezione di singole parole in una conversazione viene chiamata **speech segmentation** (=segmentazione del parlato).

Il fatto che di solito non ci siano spazi (percepiti) tra le parole diventa evidente quando ascoltiamo qualcuno parlare una lingua straniera. A una persona che non è familiare con quella lingua, le parole sembrano susseguirsi in una stringa ininterrotta. Tuttavia, a colui che parla quella lingua, le parole sembrano essere separate, così come le parole della tua lingua madre ti sembrano essere separate. In qualche modo risolviamo il problema della speech segmentation e dividiamo il flusso continuo del segnale acustico in una serie di singole parole.

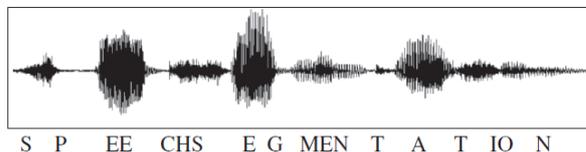


Figure 13.12 Sound energy for the words “speech segmentation.” Notice that it is difficult to tell from this record where one word ends and the other begins. Speech signal courtesy of Lisa Sanders.

Il fatto che possiamo percepire le singole parole in una conversazione, nonostante non ci siano pause nel segnale vocale, significa che la nostra percezione delle parole non si basa solamente sull’energia che stimola i recettori. Una cosa che ci aiuta a capire quando una parola finisce e ne inizia un’altra è la conoscenza dei significati delle parole. La relazione tra speech segmentation e significato è illustrata nella dimostrazione che segue.

### DEMONSTRATION

#### Organizing Strings of Sounds

Read the following words: Anna Mary Candy Lights Since Imp Pulp Lay Things. Now that you’ve read the words, what do they mean?

Se pensi che questa sia una lista di parole sconnesse che inizia con i nomi di due donne, Anna e Mary, non ti sbagli; ma prova a leggere questa successione di parole ad alta voce parlando rapidamente e ignorando gli spazi tra le parole sulla pagina. Quando fai questo, riesci a sentire una frase connessa che non inizia con i nomi Anna e Mary? (Per la risposta, vai a pagina 333-ma non sbirciare fino a quando non hai provato a leggere le parole rapidamente.)

Se sei riuscito a creare una nuova frase da quella successione di parole, l’hai fatto cambiando l’organizzazione percettiva dei suoni, e questo risultato è stato ottenuto grazie alla tua conoscenza del significato di quei suoni. Così come l’organizzazione percettiva della scena della foresta in Figura 5.32 dipendeva dal fatto di vedere le rocce come pattern

significativi (facce), la tua percezione di una nuova frase dipende dal fatto di conoscere i significati dei suoni che hai creato ripetendo rapidamente le parole.

Un altro esempio di come il significato e la conoscenza o esperienza pregressa siano responsabili dell'organizzazione dei suoni in parole è dato da queste due frasi:

Jamie's mother said, "Be a *big girl* and eat your vegetables."

The thing *Big Earl* loved most in the world was his car.

"Big girl" e "Big Earl" sono entrambe pronunciate allo stesso modo, quindi "sentirle" in maniera diversa dipende dal significato generale della frase nella quale queste parole si trovano.

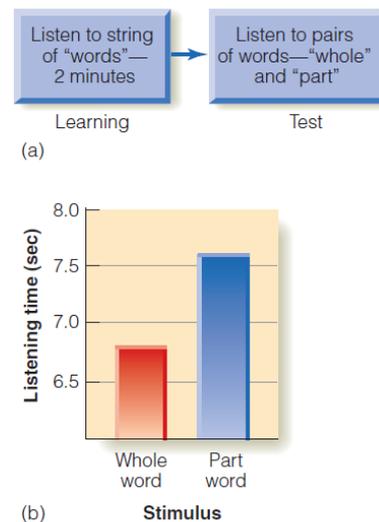
Mentre la segmentazione è aiutata dal fatto di conoscere i significati delle parole e servirsi del contesto nel quale queste parole appaiono, gli ascoltatori utilizzano anche altre informazioni per "raggiungere" la segmentazione. Quando apprendiamo una lingua, impariamo che certi suoni hanno una maggiore probabilità di susseguirsi l'un l'altro all'interno di una parola, e che altri suoni hanno una maggiore probabilità di essere separati dallo spazio tra due parole. Ad esempio, prendi in considerazione le parole *pretty baby*. In Inglese è probabile che *pre* e *ty* facciano parte della stessa parola (*pre-ty*) e che *ty* e *ba* siano separate da uno spazio in modo da far parte di due parole differenti (*pretty baby*). Così, è più probabile che lo spazio nella frase *prettybaby* si trovi tra *pretty* e *baby*.

Gli psicologi descrivono il modo in cui i suoni si susseguono l'un l'altro in una lingua in termini di **transitional probabilities** (=probabilità di transizione) – la possibilità che un suono segua un altro suono. Ogni lingua ha le proprie transitional probabilities per suoni differenti, e a mano a mano che impariamo una lingua, non solo impariamo a dire e comprendere parole e frasi, ma impariamo anche le transitional probabilities in quella lingua. Il processo di apprendimento delle transitional probabilities e di altre caratteristiche della lingua si chiama **statistical learning** (=apprendimento statistico). La ricerca ha dimostrato che bambini dell'età di 8 mesi sono in grado di mettere in atto l'apprendimento statistico.

Jennifer Saffran e colleghi (1996) hanno effettuato un primo esperimento che dimostrava la presenza di apprendimento statistico in bambini piccoli. La Figura 13.13a mostra il disegno di questo esperimento. Durante la fase di apprendimento dell'esperimento, i bambini ascoltavano quattro "parole" prive di senso come *bidaku*, *padoti*, *golabu*, e *tupiro*, che erano state combinate in ordine random in modo da creare un suono continuo di 2 minuti. Un esempio di una parte

di stringa creata dalla combinazione di queste parole è *bidakupadotigolabutupiropadotibidaku...* In questa stringa, ogni altra parola è evidenziata in neretto per aiutare ad estrapolare le parole. Nonostante ciò, quando i bambini ascoltavano queste stringhe, tutte le parole venivano pronunciate con la stessa intonazione, e non c'erano pause tra le parole che potessero indicare dove una parola finiva e dove la successiva iniziava.

**VL**



**Figure 13.13** (a) Experimental design of the experiment by Saffran and coworkers (1996), in which infants listened to a continuous string of nonsense syllables and were then tested to see which sounds they perceived as belonging together. (b) The results, indicating that infants listened longer to the "part-word" stimuli. © Cengage Learning

Poiché le parole erano state presentate in ordine random e senza spazi tra loro, la stringa di parole di 2-minuti che i bambini ascoltavano "appariva" come un miscuglio di suoni random. Tuttavia, c'era dell'informazione all'interno della stringa di parole sotto forma di transitional probabilities, che i bambini potevano potenzialmente utilizzare per determinare quali gruppi di suoni erano parole. Le transitional probabilities tra due sillabe che comparivano all'interno di una parola erano sempre 1.0. Ad esempio, per la parola *bidaku*, quando */bi/* veniva presentata, */da/* la seguiva sempre. In maniera simile, quando */da/* veniva presentata, */ku/* la seguiva sempre. In altre parole, questi tre suoni comparivano sempre insieme e nello stesso ordine, per formare la parola *bidaku*. Tuttavia, le transitional probabilities tra la *fine* di una parola e l'*inizio* di un'altra erano soltanto 0.33. Ad esempio, c'era una probabilità del 33-per cento che l'ultimo suono, */ku/* di *bidaku*, fosse seguito dal primo suono, */pa/*, di *padoti*, una probabilità del 33-per cento che fosse seguito da */tu/* di *tupiro*, e una probabilità del 33-per cento che fosse seguito da */go/* di *golabu*. Se i bambini testati da Saffran fossero stati sensibili alle transitional probabilities, avrebbero percepito stimoli

come *bidaku* o *padoti* come parole, perché le tre sillabe in queste parole sono collegate da delle transitional probabilities di 1.0. Al contrario, stimoli come *tibida* (la fine di *padoti* più l'inizio di *bidaku*) non sarebbero stati percepiti come parole, perché le componenti non erano collegate.

Per determinare se i bambini percepivano, per l'appunto, stimoli come *bidaku* e *padoti* come parole, i bambini venivano testati attraverso la presentazione di stimoli sotto forma di coppie di tre sillabe. Uno degli stimoli era una "parola" che era stata presentata in precedenza, come per esempio *padoti*. Questo era il test stimolo "whole-word" (=parola-intera). L'altro stimolo veniva creato dalla fine di una parola più l'inizio di un'altra, come per esempio *tibida*. Questo era il test stimolo "part-word" (=parola-parziale).

L'ipotesi era che i bambini avrebbero scelto di ascoltare i test stimoli part-word per più tempo rispetto ai test stimoli whole-word. Questa predizione si basava su una precedente ricerca che aveva mostrato che i bambini tendono a perdere interesse nei confronti di stimoli che vengono ripetuti.

La previsione era che i neonati avrebbero scelto di ascoltare più a lungo, durante il test, gli stimoli che presentavano le forme contratte delle parole piuttosto che gli stimoli con le parole a forma estesa. Questa supposizione si basa su studi precedenti che mostravano come i neonati abbiano la tendenza di perdere interesse per gli stimoli che vengono ripetuti e quindi diventano familiari, prestando più attenzione a nuovi stimoli di cui non hanno mai fatto esperienza. Pertanto se il bimbo percepisce lo stimolo della parola non abbreviata, come quello che è stato ripetuto più volte durante i due minuti della sessione di apprendimento, dovrebbe prestarvi meno attenzione rispetto allo stimolo più nuovo espresso come parola contratta che non percepisce come una parola esistente. Saffran misurò per quanto tempo il bambino ascoltava ogni singolo suono presentando una luce lampeggiante vicino allo speaker da dove proveniva il suono. Quando la luce attraeva l'attenzione del bambino il suono iniziava e persisteva finché il piccolo non distoglieva lo sguardo dalla luce. Dunque, in questo modo, è stato possibile controllare per quanto tempo i bambini ascoltassero ogni singolo suono, sulla base di quanto tempo avessero passato a fissare la fonte luminosa. La figura 13.13b mostra che i neonati facevano, come predetto, ascolti più lunghi se si trattava di stimoli consistenti in parole abbreviate. Questi risultati sono impressionanti soprattutto perché i bambini non avevano mai sentito prima le parole presentate, non hanno avuto pause tra l'ascolto di una parola e l'altra e hanno ascoltato le serie di parole solamente per due minuti. Da risultati come questi, possiamo concludere che utilizzare la probabilità di transizione per passare da una condizione con suoni segmentati ad una con parole, inizi, in età precoce.

## Prendere in considerazione le caratteristiche dello speaker

Quando stai prendendo parte ad una conversazione, ascoltando una lezione o un dialogo in un film, di solito ti focalizzi su determinare il significato di ciò che viene detto. Mentre stai capendo questi messaggi, stai anche, forse senza rendertene conto, assimilando le caratteristiche dello speaker. Queste caratteristiche dette caratteristiche indicali trasmettono informazioni sul relatore come la sua età, il genere sessuale, il Paese d'origine, lo stato emotivo e quando esso sia sarcastico o serio. Considera per esempio la seguente barzelletta:

*un giorno un professore di linguistica stava tenendo una lezione alla sua classe. "in inglese"- disse- "una doppia negazione viene percepita come un'affermazione, in alcune lingue tuttavia, come il russo una doppia negazione, esprime una frase negativa. Dunque non c'è una lingua in cui la doppia affermazione possa formare una negazione." A quel punto, si levò una voce dal fondo della stanza esclamando "sì, certo!"*

questa barzelletta è umoristica perché "sì, certo" accosta due parole positive che, malgrado l'affermazione del professore, produce un'asserzione negativa che la maggior parte delle persone che ne conoscono l'uso nell'inglese contemporaneo interpreterebbero come "io non sono d'accordo". Il punto di questo esempio non è soltanto che "sì certo" possa voler dire "io non sono d'accordo" ma che il significato di questa frase sia determinato dalla nostra conoscenza dell'utilizzo dell'inglese corrente e anche (se noi avessimo veramente ascoltato l'osservazione dello studente) dal tono di voce dello speaker che in questo caso sarebbe stato molto sarcastico.

Il tono dell'oratore è uno dei fattori che aiuta gli ascoltatori a determinare il significato di ciò che viene detto. La maggior parte delle ricerche sulle caratteristiche indicali si è concentrata su come la percezione del discorso sia influenzata dall'identità dello speaker. Thomas Palmeri, Stephen Goldinger, e David Pisoni (1993) hanno dimostrato quanto possa essere forte sulla percezione l'effetto dovuto all'identità di chi parla presentando agli ascoltatori una sequenza di parole. Dopo ogni parola, gli ascoltatori indicavano se il termine fosse una parola nuova (se era la prima volta che appariva) o una vecchia (se era apparsa precedentemente nella sequenza).

Hanno scoperto che gli ascoltatori reagivano più rapidamente ed erano più accurati quando lo stesso speaker diceva tutte le parole rispetto a quando speaker diversi pronunciavano parole diverse. Ciò significa che gli ascoltatori comprendono in due livelli di informazione del termine: (1) il significato e (2) le caratteristiche della voce dello speaker. Dai risultati di questo esperimento e dagli altri che abbiamo discusso possiamo concludere che la percezione del discorso dipenda sia dalle informazioni che vanno dal basso verso l'alto fornite dal segnale acustico, sia dalle informazioni dall'alto verso il basso fornite dai significati delle parole, le frasi, le conoscenze grammaticali dell'ascoltatore e le informazioni che possiede sulle caratteristiche della voce dello speaker.

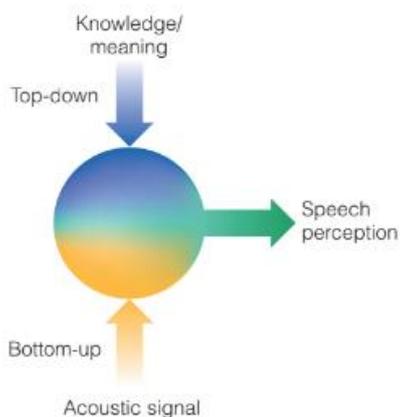
### (Figura 13.14)

Possiamo cogliere l'interazione del segnale acustico finalizzato al discorso e il significato del discorso quando realizziamo che nonostante noi utilizziamo il significato per aiutarci a capire il segnale acustico, il segnale acustico è il punto di partenza per determinare il significato.

Vediamola in questo modo: potrebbero esserci abbastanza informazioni nella mia grafia disordinata così che una persona che stia usando il sistema di elaborazione dal basso verso l'alto può decifrarla solamente sulla base degli scarabocchi sulla pagina, ma la mia scrittura è molto più facile da decifrare quando, si usa l'elaborazione dall'alto verso il basso, in questo modo la persona interpreta il significato delle parole all'interno dello scritto. Come nelle precedenti esperienze nell'ascoltare la voce di una persona

particolare rende più facile capire quella persona in un secondo momento, se precedentemente avessi letto la mia scrittura sarebbe più facile decifrare gli scarabocchi nella pagina.

La percezione del discorso funziona apparentemente in modo simile. La maggior parte delle informazioni è contenuta nel segnale acustico, interpretare il significato e le caratteristiche indicali all'interno del racconto rende più semplice la comprensione del discorso.



**Figura 13.14** – La percezione del discorso è il risultato dell'elaborazione dall'alto verso il basso (basata su conoscenza e significato) e dell'elaborazione dal basso verso l'alto (basata sul segnale acustico) che lavorano assieme.

## La percezione del discorso e il cervello

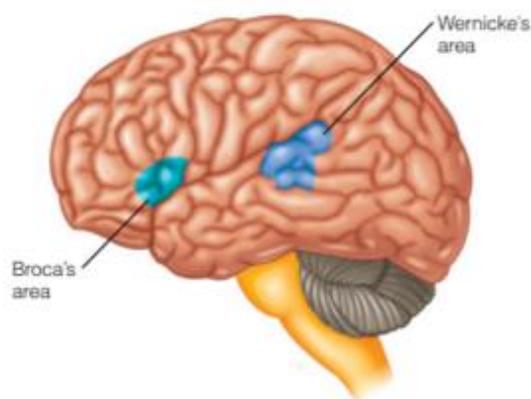
La ricerca delle basi fisiologiche per la percezione del discorso risale almeno al diciannovesimo secolo ma progressi considerevoli sono stati fatti solo recentemente nella comprensione delle basi psicologiche della percezione del discorso e il riconoscimento delle parole dette.

### Localizzazione corticale della percezione del discorso

Basandosi nei loro studi sui pazienti con danni cerebrali i ricercatori del diciannovesimo secolo Broca e Wernicke mostrano che un danno a specifiche aree del cervello causa problemi al linguaggio, detti afasie (**figura 13.15**). Ci sono numerose forme di afasie, con i sintomi specifici che dipendono dall'area danneggiata e dall'entità del danno. I pazienti con danni nell'area di Broca presentano nel lobo frontale una condizione detta afasia di Broca. Questi pazienti hanno elaborato uno stile linguistico elaborato e innaturale e

possono parlare solamente utilizzando frasi sintatticamente corte e semplici. Le persone affette da questo tipo di afasia sono comunque, in qualche modo, capaci di comprendere cosa gli altri stiano dicendo.

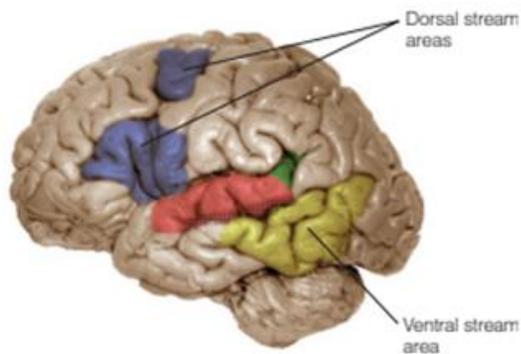
I pazienti con danni all'area di Wernicke, nel lobo temporale riportano l'afasia di Wernicke. Questi pazienti possono parlare fluentemente ma quello che dicono è estremamente disorganizzato e privo di significato. Le persone con l'afasia di Wernicke hanno gravi difficoltà a capire cosa gli venga detto dalle altre persone. Nella sua forma più grave questo tipo di afasia può portare ad una condizione detta "sordità di parola" nella quale il paziente non può riconoscere le parole, pur mantenendo intatta l'abilità di ascoltare toni puri (Kolb e Whishaw, 2003).



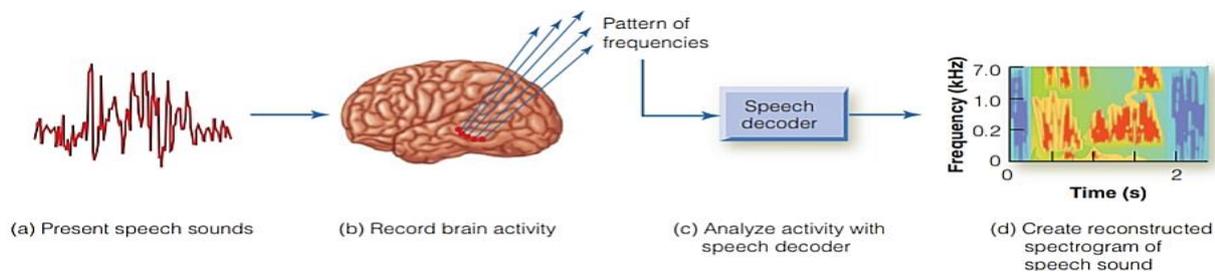
**Figura 13.15** le aree di Broca e Wernicke sono state identificate in ricerche recenti come specializzate per la produzione e comprensione del linguaggio.

La ricerca moderna è andata oltre la localizzazione della produzione del discorso e la sua percezione in queste due aree, attraverso studi successivi su pazienti con danni cerebrali usando l'*imaging cerebrale* (vedi metodo: doppia dissociazione in neuropsicologia, capitolo 4, pagina 85) per localizzare le aree del cervello connesse all'uso del linguaggio. Un esempio di scoperta in neuropsicologia è che alcuni pazienti con danni al lobo parietale hanno difficoltà nello discriminare tra sillabe (Blumstein et al. 1977 - Damasio e Damasio, 1980). Nonostante potremmo aspettarci che la difficoltà nel discriminare tra sillabe renda difficile capire le parole, alcuni pazienti che hanno difficoltà a discriminare le sillabe possono comunque capire il significato delle parole (Micelli et altri, 1980). Risultati come questi illustrano la complessa relazione tra il funzionamento del cervello e la percezione del linguaggio. Misurare l'attività del cervello ha prodotto risultati più chiari, per esempio Pascal Belin e i suoi collaboratori (2000) usarono l' fMRI per localizzare

“un’area della voce” nel solco temporale superiore umano (STS) che è attivata maggiormente da voci umane rispetto che da altri suoni, Catherine Perrodin e collaboratori (2011) documentarono dei neuroni nel lobo temporale della scimmia, che hanno chiamato “cellule voce” perché rispondevano più vigorosamente alle registrazioni di richiami di scimmie che ai richiami di altri animali o a suoni “non voce”. L’area della voce e le cellule voce sono localizzate nel lobo temporale che è parte del sistema di elaborazione del “*che cosa*” come abbiamo descritto nel capitolo 12 (vedi pagina 300). Nel descrivere l’organizzazione corticale come detto nel capitolo 12, abbiamo visto che il percorso del cosa è coinvolto nella localizzazione dei suoni (figura 12.17). A proposito di questa idea del modello a doppio flusso per quanto riguarda l’udito, esiste un modello a doppio flusso di percezione del discorso, questo processo consiste in una via ventrale (o che cosa) che inizia nel lobo temporale ed è responsabile del riconoscimento delle parole e una via dorsale (o dove) che inizia nel lobo parietale che è responsabile del collegamento del segnale acustico ai movimenti usati per produrre i suoni e le parole all’interno dei discorsi (**Figura 13.16**, Hickock e Poppel,2007).



**Figura 13.16** il modello del doppio flusso della percezione del linguaggio propone una via ventrale che è responsabile del riconoscimento del linguaggio e una via dorsale che collega il segnale acustico ai movimenti motori. Le aree blu sono associate con la via dorsale e le aree gialle con la via ventrale. Anche le aree rosse e verdi sono coinvolte nell’analisi degli stimoli linguistici.



a) suoni del linguaggio presenti	b) registrazione attività cerebrale
----------------------------------	-------------------------------------

c) Analisi attività con decodificatore di linguaggio	d) Creare uno spettrogramma ricostruito del suono della voce
--	--

**Figura 13.17** Pasley e altri (2012) procedura per creare spettrogrammi ricostruiti da segnali elettrici registrati dal cervello. Vedi il testo per i dettagli.

Questo è simile alla situazione che abbiamo descritto nel capitolo 5 per la percezione di volti (pagina 120). Abbiamo visto che la percezione di volti comporta diversi aspetti, comprese l'identificazione del volto, la lettura delle espressioni, notare dove il volto sta guardando, e valutare la bellezza del volto, e che i meccanismi per la percezione dei volti sono quindi distribuiti attraverso diverse aree. Ci sono inoltre un numero di diversi aspetti per parlare di percezione. È influenzata da fattori cognitivi come il significato delle parole, il contesto delle frasi, e la familiarità con la voce di un oratore; è legata alla vista; e ha una forte componente emotiva. Non è sorprendente, quindi, che la percezione del linguaggio comporti diverse aree della corteccia.

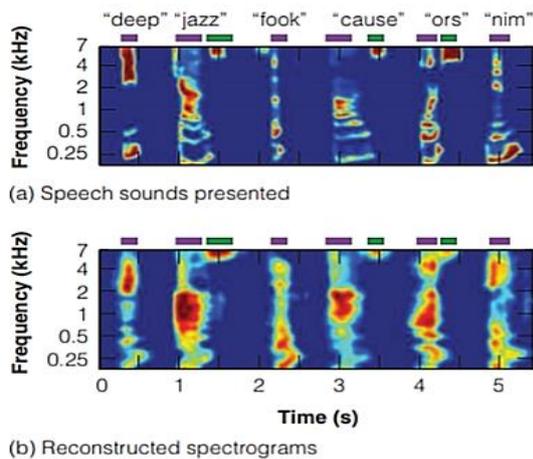
## Ricostruzione del Linguaggio Dal Segnale della Corteccia

La ricerca sopra descritta si occupa di dove il linguaggio viene processato nella corteccia. Un altro modo di approcciare la psicologia della percezione del linguaggio è di chiedersi come lo schema dei segnali elettrici delle aree del linguaggio rappresenta i suoni del discorso.

Brian Pasley e collaboratori (2012) hanno approcciato questa domanda creando un "decodificatore vocale" analogo al visivo "decodificatore di scene" che abbiamo descritto nel capitolo 5, in grado di determinare, basandosi sull'analisi dei segnali fMRI, il tipo di scena che una persona stava osservando (Naselaris e altri, 2009; vedi "Leggere il Cervello", Capitolo 5, pagina 116). Per sviluppare un decodificatore vocale, Pasley non usò i segnali fMRI ma segnali registrati da elettrodi posti sulla superficie del cervello. Per realizzarlo, si affida all'aiuto di pazienti che

stavano aspettando di avere un'operazione al cervello per eliminare gravi crisi epilettiche. Per determinare quali aree del cervello sarebbero dovute essere rimosse per eliminare le crisi, i neurochirurghi posizionavano gli elettrodi sulla superficie del cervello e registravano l'attività di una settimana.

Come mostrato in **Figura 13.17**, Pasley presentò i suoni del linguaggio a questi pazienti (**Figura 13.17a**) e registrò l'attività captata dall'elettrodo vettore (**Figura 13.17b**). Perché diverse frequenze sono rappresentate da attività in diversi luoghi della corteccia uditiva (riguardare indietro alla Figura 11.31, pagina 281), Pasley riuscì a determinare lo schema di frequenze negli stimoli del linguaggio dall'attività registrata da diversi posizionamenti degli elettrodi. Questo schema di attività fu poi analizzato con il decodificatore vocale (**Figura 13.17c**), che creò uno spettrogramma vocale del suono che veniva presentato (**Figura 13.17d**). Questo spettrogramma è chiamato spettrogramma "ricomposto" perché è costruito dai segnali elettrici creati dal decodificatore vocale come mostrato nella riga sotto (**Figura 13.18b**). La corrispondenza tra gli spettrogrammi effettiva e ricostruita è lontana dall'essere perfetta, ma quando un dispositivo "playback" è usato per convertire questi schemi di frequenza in suoni, è possibile sentire suoni di linguaggio che, in diversi casi, possono essere riconosciuti come le parole che il paziente aveva udito.



**Figura 13.18** (a) Spettrogrammi di suoni di linguaggio presentati ai soggetti di Pasley. (b) Spettrogrammi vocali costruiti con decodificatori vocali dai segnali elettrici registrati nella corteccia uditiva.

Questo risultato è importante perché ci porta più vicini a scoprire come il linguaggio è rappresentato nel cervello umano. Inoltre, riuscire a determinare il linguaggio dall'attività cerebrale potrebbe potenzialmente aiutare le persone che non possono parlare a causa di patologie come ALS (sclerosi laterale amiotrofica; malattia di Lou Gehrig), che paralizza il corpo mentre lascia intatta la mente. Perché pensare di parlare può creare segnali cerebrali simili a quelli che si verificano durante un discorso reale, è possibile che un giorno diventi possibile per le persone con ALS comunicare avendo dei propri pensieri sul discorso trasformandoli in suoni usando un decodificatore vocale simile a quello descritto qui.

## QUALCOSA DA CONSIDERARE:

# Linguaggio percepito e Azione

Un'importante caratteristica del linguaggio è che non lo percepiamo solamente, lo produciamo anche. Questa stretta connessione tra percepire il linguaggio e produrlo condusse Alvin Liberman ed i collaboratori (1963,1967) a proporre una teoria chiamata **teoria motoria della percezione del linguaggio**. Questa teoria propose che (1) sentire un particolare linguaggio sonoro attivasse il meccanismo motorio controllante le articolazioni, come la lingua e le labbra, responsabili della produzione dei suoni; e (2) l'attivazione di questi meccanismi motori, a loro volta, attivasse meccanismi aggiuntivi che ci permettono di *percepire* il suono. Quindi, la teoria motoria propone che l'attività di meccanismi motori è il primo gradino verso la percezione del linguaggio.

Quando la teoria motoria fu proposta la prima volta negli anni 1960, fu molto discussa. Nei decenni che seguirono, la teoria favorì un gran numero di esperimenti, qualcuno ottenne risultati che supportarono la teoria, ma molti ottennero risultati che la discutevano. È difficile per la teoria motoria spiegare, per esempio, come persone con il cervello danneggiato che inabilita il loro sistema motorio linguistico possano ancora percepire il linguaggio (Lotto e altri., 2009). Prove come questa condussero alle odierne ricerche sul linguaggio che rifiutano ampiamente l'idea che la nostra percezione del linguaggio è basata sull'attivazione di meccanismi motori.

Sebbene le prove discutano l'idea che l'attivazione di meccanismi motori sia necessaria per la percezione del linguaggio, ci sono prove di *collegamenti* tra meccanismi motori e percezione del linguaggio. Uno dei risultati che supporta quest'idea è la scoperta dei neuroni specchio. Nel Capitolo 7, abbiamo visto che i neuroni specchio nelle scimmie rispondono sia quando la scimmia compie un'azione e quando la scimmia vede qualcun altro compiere l'azione. Un tipo di neuroni specchio connessi con l'ascolto sono chiamati *neurone specchio audiovisivo*. Questi neuroni si attivano quando una scimmia *compie un'azione che produce un suono* (come rompere una nocciolina) e quando la scimmia *sente il suono* (il suono di una nocciolina rotta) che risulta dall'azione (Kohler, 2002; vedi Capitolo 7, pagina 167). Curiosamente, i neuroni specchio che sono stati studiati nella scimmia si trovano in un'area quasi equivalente nell'area di Broca negli umani; per questa ragione, alcuni ricercatori hanno proposto uno stretto collegamento tra i neuroni specchio e il linguaggio (Arbib, 2001).

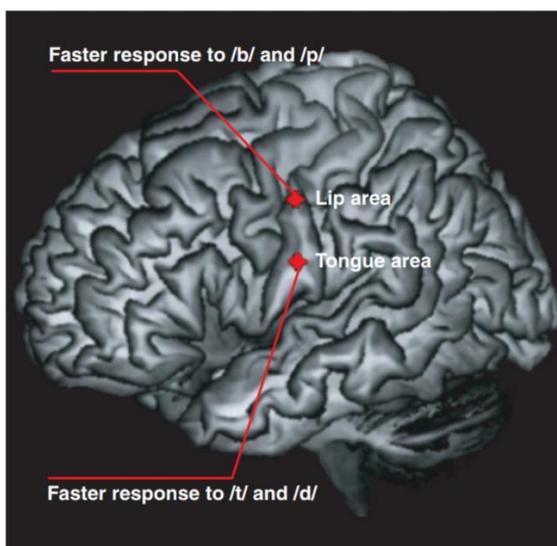
Ma non c'è nessuna prova che collega la *percezione* del linguaggio e la produzione del linguaggio negli umani? Alessandro D'Assilio e colleghi (2009) dimostrarono un collegamento tra produzione e percezione mostrando che aumentando l'attivazione delle aree motrici collegate producendo suoni come /b/ e /p/, che comportano l'articolazione labiale (contraendo le labbra) aiuta nella percezione di questi suoni. Analogamente, la stimolazione di un'area motrice collegata producendo suoni come /t/ e /d/, che comportano un'articolazione dentale (la lingua si contrae dietro i denti) aiuta nella percezione di questi suoni.

Il compito dei soggetti nell'esperimento di D'Ausilio era di premere un bottone il più velocemente possibile per indicare quale suono sentivano in ogni prova. Nella condizione base, i soggetti portavano a termine questo compito senza alcuna stimolazione del cervello.

Nella condizione di stimolazione, brevi pulsazioni di *stimolazioni magnetiche transcraniche focali*, le quali possono stimolare una piccola area bersagliata del cervello, furono presentate appena prima che il soggetto sentisse il suono, o nell'area della corteccia motrice responsabile di creare l'articolazione labiale

(labbra) o l'area che crea l'articolazione dentale (lingua e denti). (Vedi Metodo: stimolazione magnetica transcranica (TMS), Capitolo 8, pagina 192.)

La **Figura 13.19** mostra i siti di stimolazione nell'area motrice della corteccia. La stimolazione dell'area labiale risultò reagire più velocemente nei fonemi labiali (/b/ e /p/) e la stimolazione dell'area della lingua risultò reagire più velocemente ai fonemi dentali (/t/ e /d/). Basato su questi risultati, D'Ausilio suggerì che questa attività nella corteccia motrice può influenzare la percezione del linguaggio. Ricerche come questa, insieme a molti altri studi, mostrarono che lo stretto collegamento tra l'attività motoria e la percezione reggeva non solo per la percezione visiva (vedi Capitolo 7) ma anche per la percezione del linguaggio.



**Figura 13.19** Siti della stimolazione magnetica transcranica dell'area motrice per labbra e lingua. La stimolazione dell'area labiale aumenta la velocità di risposta a /b/ e /p/. La stimolazione dell'area della lingua aumenta la velocità di risposta a /t/ e /d/.

ASPETTO DELLO SVILUPPO EVOLUTIVO: LA PERCEZIONE DELL'INFANTE DEL DISCORSO

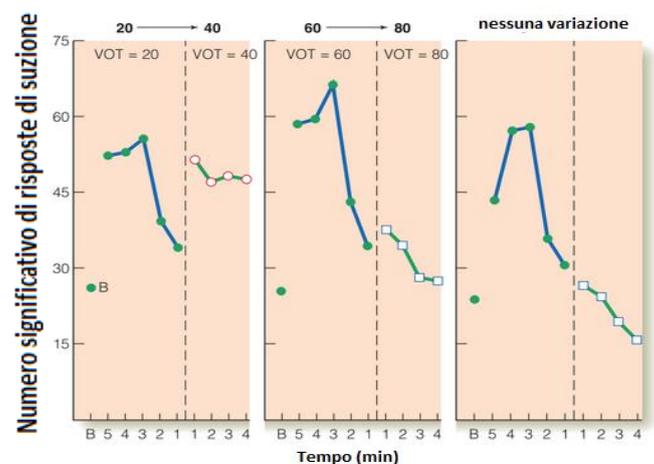
Abbiamo visto dagli esperimenti di Saffran (**figura 13.13**) che gli infanti possono usare dati del discorso per realizzare la segmentazione di quest'ultimo. Inoltre, la ricerca ha mostrato una percezione categoriale nei bambini di 1 anno.

La percezione categoriale dei fonemi

La percezione categoriale è stata dapprima riportata negli adulti nel 1967 (Liberman et al., 1967). Nel 1971, Peter Eimas e collaboratori dettero il via alla moderna era di ricerca sulla percezione del discorso nei bambini. Negli esperimenti di percezione categoriale hanno usato la procedura dell'abituazione per mostrare che infanti di 1 mese si comportano in modo simile agli adulti. Alla base di questi esperimenti c'era l'osservazione che i bambini succhieranno il capezzolo al fine di sentire una serie di corti dialoghi di suoni, tuttavia siccome gli stessi suoni vengono ripetuti, il succhiare del bambino alla fine lo abitua ad un livello basso. Presentando un nuovo stimolo dopo che la frequenza di suzione è diminuita, Eimas ha determinato se il neonato ha percepito il nuovo stimolo come se fosse lo stesso suono oppure uno diverso da quello precedente.

I risultati dell'esperimento di Eimas e collaboratori sono illustrati nella **figura 13.20**. Il numero di risposte di suzione quando non viene presentato nessun suono è indicato dal punto B. Quando un suono con un tempo di attacco della sonorità (VOT) di 20ms (suoni come "ba" di un adulto) è presentato mentre l'infante sta succhiando, la suzione aumenta fino ad un livello alto e poi inizia a decrescere. Quando il VOT viene variato a 40ms (linea tratteggiata; suoni come "pa" di un adulto), la suzione cresce, come indicato dai punti a destra della linea tratteggiata. Questo significa che il bambino percepisce una differenza tra i suoni con VOT di 20 e 40ms. Il grafico centrale, peraltro, mostra che la variazione del VOT da 60 a 80ms (entrambi suoni come "pa" di un adulto) ha soltanto un piccolo effetto sulla suzione, indicando che i bambini percepiscono in minima parte, sempre che percepiscano qualcosa, le differenze tra i due suoni. Infine, i risultati di un gruppo di controllo (grafico di destra) mostrano che quando il suono non viene variato, il numero di risposte di suzione decresce durante l'esperimento.

**Figura 13.20** Risultati di un esperimento sulla percezione



categoriale nei bambini usando la procedura dell'abituazione. Nel pannello a sinistra, il VOT viene variato da 20 a 40ms (attraverso il limite fonetico). Nel pannello centrale, il VOT viene cambiato da 60 a 80ms (non attraversando il limite fonetico). Nel pannello di destra, il VOT non viene variato. Vedi il testo per i dettagli.

Questi risultati mostrano che quando il VOT viene spostato attraverso il limite fonetico di un adulto medio (grafico a sinistra), i bambini percepiscono un cambiamento nel suono, e quando il VOT viene variato dalla stessa parte del limite fonetico (grafico centrale), i neonati percepiscono un minimo o nessun cambiamento. Il fatto che gli infanti di 1 mese siano capaci della percezione categoriale è particolarmente impressionante poiché questi non hanno praticamente nessuna esperienza nel produrre suoni di un discorso e soltanto una limitata esperienza nell'ascoltarli. Tuttavia la storia riguardo la percezione di fonemi si estende oltre la scoperta che bambini molto piccoli possono percepire fonemi. Come vedremo nella prossima parte, l'abilità di percepire fonemi è influenzata dalla lingua che un neonato sente durante il primo anno di vita.

## Imparare i suoni di una lingua

Se l'inglese è la tua lingua madre, potresti aver notato che molte persone la cui lingua principale è il giapponese pronunciano parole che iniziano in "r" con un suono "l", perciò diranno "lent (prestato)" invece di "rent (affittare)". Questa difficoltà nel produrre alcuni fonemi è collegata ad un'incapacità di percepire questi fonemi, come indicato dalla constatazione che i madrelingua giapponesi hanno difficoltà nel distinguere tra le parole *lent* e *rent*.

Ciò che rende questa difficoltà particolarmente interessante è che i bambini piccoli in tutte le culture possono indicare la differenza tra i fonemi che creano tutti i suoni dei discorsi usati nelle varie lingue del mondo, tuttavia dall'età di 1 anno, perdono questa abilità di distinguere tra alcuni di questi suoni (Kuhl, 2000). Pertanto, i bambini giapponesi di 6 mesi possono riconoscere una differenza tra la /r/ e la /l/ usate nell'inglese americano tanto bene quanto i bambini americani. Però, a 12 mesi, i neonati giapponesi confondono /r/ e /l/. In quello stesso periodo, i bambini americani migliorano nell'identificare la differenza tra questi due suoni (Kuhl et al., 1997; Strang, 1995).

Perché si verifica questo cambiamento? La risposta a questa domanda implica la *plasticità esperienza dipendente*, un cambiamento nell'abilità del cervello di rispondere ad uno specifico stimolo che si verifica a risultato di un'esperienza. Abbiamo introdotto la *plasticità esperienza dipendente* nel capitolo 3 quando abbiamo descritto come gattini cresciuti in un ambiente fatto interamente da linee verticali porta il cervello dei gattini a contenere neuroni che rispondono solo a linee verticali (pagina 68).

Una prova, a supporto dell'idea che il cambiamento nella percezione del linguaggio che avviene ogni tanto dopo i 6 mesi probabilmente coinvolga la plasticità esperienza dipendente, è stata fornita da Maritza Rivera-Gaxiola e collaboratori (2005). Essi hanno registrato i potenziali elettrici dalla superficie della corteccia di infanti americani di 7 e 11 mesi provenienti da famiglie di lingua inglese, in risposta a coppie di suoni che sembrano gli stessi ad un adulto madrelingua inglese, ma che sono percepiti diversamente da adulti di lingua spagnola. A 7 mesi, la risposta elettrica a questi due suoni era diversa nei bambini madrelingua inglese, però a 11 mesi il responso è diventato lo stesso.

I risultati forniscono un parallelo fisiologico all'esperienza dei bambini piccoli giapponesi descritta prima. Una coppia di suoni può essere percepita come differente oppure può causare diverse risposte fisiologiche ad una giovane età, tuttavia se il bambino non ha esperienza nel discriminare tra i 2 suoni, allora il bambino perderà l'abilità di riconoscerne la differenza, e le risposte fisiologiche ai suoni diventeranno le stesse. Evidentemente, il cervello è modellato dall'esperienza per rispondere a suoni che sono utilizzati nella specifica lingua che il bambino sta imparando.