



UNIVERSITÀ DEGLI STUDI DI MESSINA

DIPARTIMENTO DI SCIENZE COGNITIVE, PSICOLOGICHE,
PEDAGOGICHE E DEGLI STUDI CULTURALI

DOTTORATO DI RICERCA IN SCIENZE COGNITIVE

XXXIII CICLO

IL CORPO DELLA VOCE

**Integrazione audiovisiva e gesto articolatorio nello sviluppo del
linguaggio e nell'apprendimento di una seconda lingua**

Dottoranda:

Marta ARCOVITO

Matricola: 479673

Supervisore tesi:

Chiar.ma Prof.ssa Alessandra Maria FALZONE

Coordinatore del dottorato:

Chiar.ma Prof.ssa Alessandra Maria FALZONE

Anno Accademico 2019/2020

Indice

Introduzione.....	4
MULTIMODALITÀ	11
1.1 Il linguaggio umano: un sistema di sistemi.....	11
1.2 Meccanismi di acquisizione del linguaggio	16
1.3 La percezione del linguaggio verbale durante l'infanzia: un processo multisensoriale.....	30
IL CORPO COME CONTESTO PERCETTIVO: IL GESTO ARTICOLATORIO DURANTE L'INFANZIA... 33	
2.1 L'integrazione audiovisiva nei neonati	33
2.2 Il ruolo del gesto articolatorio nel processamento dei suoni linguistici.....	38
2.3 Bilingui vs monolingui: riconoscimento visivo del linguaggio verbale	41
2.4 Attenzione selettiva alla regione della bocca	44
IL CORPO COME CONTESTO PERCETTIVO: IL GESTO ARTICOLATORIO DOPO L'INFANZIA	47
3.1 Gesto articolatorio e processamento del linguaggio verbale.....	47
3.2 Il ruolo del gesto articolatorio in età adulta	50
3.3 Lettura del labiale e integrazione audiovisiva	51
SISTEMA ARTICOLATORIO E MULTIMODALITÀ INTRINSECA	58
4.1 Il sistema motorio del linguaggio verbale	58
4.2 Le componenti motorie della percezione del linguaggio verbale.....	62
4.3 La preferenza multimodale per il maternese.....	70
CENTRALITÀ DELLA MULTIMODALITÀ DEL LINGUAGGIO NELL'APPRENDIMENTO DI UNA SECONDA LINGUA	75
5.1 Il processo di apprendimento di una seconda lingua	75
5.2 La didattica della seconda lingua	84
5.3 Gesto articolatorio e seconda lingua.....	90
CONCLUSIONI	98
6.1 La doppia natura multimodale del linguaggio verbale.....	98
6.2 La multimodalità del linguaggio verbale e l'apprendimento di una seconda lingua	101
6.3 Applicazioni pratiche.....	110
Riferimenti bibliografici	120

Introduzione

È quando una delle componenti di un sistema non è più accessibile, che ci si rende conto dell'impatto, prima poco visibile nell'uso fisiologico, di quella componente sul funzionamento del sistema, anche se quest'ultimo continua comunque ad assolvere alla sua funzione.

Questa considerazione è applicabile, per esempio, alla condizione in cui interagiamo con una persona che indossa una mascherina a protezione delle vie respiratorie e non abbiamo, quindi, accesso alla visione della sua bocca. Quello che ci interessa è sentire, non vedere la bocca, eppure la sensazione è spesso quella di sentire di meno.

La bocca del parlante rappresenta la componente articolatoria visibile del linguaggio. Essa non è una componente necessaria per percepire il suono, ma ha un valore intrinseco legato al suono linguistico, la cui disamina può aiutare a comprendere in che modo essa permei il linguaggio verbale a un livello profondo, non solo nella funzione di produzione delle parole.

L'indagine del ruolo della componente articolatoria e multimodale del linguaggio verbale sta alla base di questo lavoro.

Il percorso che segue mira a comprendere e inquadrare la funzione del sistema articolatorio umano e la relazione che esso ha con il linguaggio verbale che produce.

Al passaggio d'aria, la vibrazione delle corde vocali produce onde sonore che si propagano poi nell'aria e arrivano all'orecchio di un'altra persona sotto forma di voce.

L'apparato di fonazione di cui è dotato, permette all'uomo di dare vita, attraverso la voce, a una delle funzioni cognitive più complesse e più studiate sia a livello filogenetico che ontogenico: il linguaggio verbale.

Le onde sonore sono onde longitudinali, ossia onde in cui le vibrazioni prodotte dalla sorgente sonora producono un'alterazione nella densità delle molecole dell'aria, che iniziano così a muoversi parallelamente alla direzione dell'onda in un alternarsi di compressioni e rarefazioni della densità molecolare, fino al ritorno al punto di equilibrio e quindi al silenzio (Fishbane et al., 1993).

Essendo l'atto verbale un atto non automatico, bensì volontario, perché possa avere origine e destinazione, l'onda sonora linguistica ha bisogno di un corpo (il cui sistema di fonazione agisce come sorgente sonora), di un mezzo di propagazione (solitamente l'aria), e di un sistema ricevente che processi ed elabori l'onda sonora.

Il linguaggio verbale è il prodotto di un fenomeno meccanico che crea un'onda sonora che altera la densità media delle molecole dell'aria. L'onda sonora, nella forma di suono linguistico, può poi essere processata dal sistema uditivo ed elaborata da quello nervoso del ricevente.

Il linguaggio verbale si configura quindi come una sorta di passaggio di oscillazioni molecolari che vengono originate da un corpo e il cui output sonoro viene processato ed elaborato da un altro corpo.

Ma cosa rimane dell'elemento corporeo che le produce, nella complessa concatenazione di onde sonore che danno vita al linguaggio verbale?

Che ruolo ha l'esperienza corporea nel processamento e nell'elaborazione di queste oscillazioni molecolari?

È possibile che le onde sonore linguistiche trasferiscano, oltre all'informazione sonora, un'informazione corporea del linguaggio verbale, propedeutica al suo processamento?

E, infine, in che modo tale eventuale informazione corporea può entrare in gioco nella percezione di input sonori poco chiari o poco familiari, come ad esempio quelli di una lingua diversa da quella materna?

L'obiettivo di questo lavoro è quello di fare luce, partendo da queste domande, sul ruolo della dimensione corporea nell'apprendimento del linguaggio verbale proprio relativamente all'atto che lo realizza: il gesto articolatorio.

Il gesto articolatorio è l'atto motorio compiuto dal tratto vocale e dai muscoli facciali grazie al quale si realizza la produzione linguistica. La parte visibile del gesto articolatorio è rappresentata dal movimento della bocca di chi parla.

Nella prima parte di questo lavoro, si approfondisce in che modo la visione del gesto articolatorio altrui e l'integrazione multisensoriale del doppio input, visivo (gesto articolatorio) e acustico (voce), influenzino la percezione sonora, approfondendo gli aspetti legati all'incidenza dell'esperienza linguistica sull'attenzione selettiva agli indizi articolatori e sull'importanza che essi acquistano nel processamento dei suoni linguistici nel corso dello sviluppo.

Il gesto articolatorio è presentato nel ruolo di elemento che attribuisce salienza percettiva al volto del parlante, arricchendo il contesto in cui è collocato e processato l'input verbale.

Questo percorso attraverso il ruolo del gesto articolatorio quale contesto percettivo che influisce sul processamento sonoro permette di spostarsi con più agilità verso un livello più profondo e più radicato del rapporto tra sistema articolatorio e linguaggio verbale.

A partire dal dato sulla minore efficienza nel processamento audiovisivo degli input verbali negli individui con storia di disturbo specifico del linguaggio (Kaganovich et al., 2016) e in quelli con diagnosi di dislessia (van Laarhoven et al., 2018), (Schaadt et al., 2019), (Rüsseler et al., 2018), si apre la seconda parte della trattazione, tesa a fornire elementi sulla natura motoria degli aspetti produttivi e percettivi del linguaggio verbale.

Oltre a una minore efficienza nel processo di integrazione audiovisiva degli input verbali (acustico e visivo), infatti, è stata riscontrata negli individui con diagnosi di dislessia una anomalia nella velocità di articolazione dei suoni linguistici: essa risulta più lenta rispetto a quella del gruppo di controllo, sia relativamente alla produzione linguistica, sia relativamente alla pianificazione motoria del gesto articolatorio durante il flusso verbale (Fawcett & Nicolson, 2002). Inoltre, buone capacità di lettura del labiale sono correlate a migliori capacità di produzione e articolazione verbale (Heikkilä et al., 2017).

Nella parte conclusiva si analizza il ruolo che la natura motoria del linguaggio verbale può avere nel contesto dell'apprendimento di una seconda lingua,

spiegando che cosa possa segnalare il ruolo della componente articolatoria rispetto alla relazione tra sistema motorio, percezione e produzione del linguaggio.

La capacità di percepire le differenze sonore tra i suoni di una lingua diversa da quella materna non è sempre scontata. I giapponesi, per esempio hanno grosse difficoltà nella distinzione sonora dei fonemi /r/ e /l/, che nella loro lingua non sono differenziati né acusticamente né produttivamente (Miyawaki et al., 1975).

Tale capacità di identificazione migliora però a seguito di un training specifico per l'identificazione del contrasto fonetico. Il miglioramento non si limita all'incremento della capacità percettiva, ma si estende al miglioramento della capacità di produzione di parole contenenti i due fonemi /r/ e /l/ (Bradlow et al., 1997).

Le indagini neurofisiologiche condotte durante il processamento dei suoni linguistici e durante la visione del solo movimento delle labbra relativo alla loro produzione rilevano l'attivazione di aree cerebrali motorie coinvolte nella produzione di linguaggio (Fadiga et al., 2002), (Watkins et al., 2003), (Wilson et al. 2004).

Tuttavia, nel caso di fonemi di una lingua straniera, l'attivazione nelle aree motorie è maggiore rispetto a quella rilevata durante l'ascolto della lingua materna (Wilson & Iacoboni, 2006) e aumenta all'aumentare della difficoltà di identificazione dei fonemi, come nel caso dei suoni consonantici /r/ e /l/ per i giapponesi (Callan, Tajima et al., 2003), (Callan et al., 2004).

L'incremento dell'attività cerebrale nelle aree deputate alla produzione linguistica in risposta al processamento di suoni difficili di una lingua che non è la propria lingua madre e durante il processamento di suoni della propria lingua materna pronunciati con un accento straniero (Callan et al., 2014) indica per gli autori di questi studi un maggiore ricorso, nel caso di un input sonoro con cui si ha poca o nessuna familiarità, ai sistemi di controllo del feedback articolatorio-uditivo.

I miglioramenti percettivi a seguito di un training mirato (Callan, Tajima et al., 2003) sono, in quest'ottica, il risultato dell'acquisizione di mappe articolatorio-uditive e uditivo-articolarie che intervengono a facilitare l'identificazione fonetica e alle quali serve fare maggiore ricorso nel caso di suoni più difficilmente riconoscibili (Callan, Tajima et al., 2003), (Callan et al., 2004).

Il fatto che l'esposizione audiovisiva allo stimolo verbale (che includa dunque la presentazione dell'input sonoro associato al relativo gesto articolatorio) nel caso di contrasti fonetici difficili o di differenze nella durata dei dittonghi di una seconda lingua dia risultati percettivi migliori rispetto all'esposizione unicamente sonora (Navarra & Soto-Faraco, 2007), (Hirata & Kelly, 2010) è messo in relazione proprio con questi ultimi studi. Il gesto articolatorio viene presentato come una chiave d'accesso ottimale per la ricostruzione di quelle mappe uditivo-articolarie utili a un più efficace processamento sonoro.

Vengono infatti individuati due livelli di multimodalità del linguaggio verbale: il primo acustico-visivo, relativo al ruolo del gesto articolatorio quale contesto percettivo per l'input sonoro; il secondo, acustico-articolarie relativo

all'incremento dell'attività cerebrale nelle aree motorie del linguaggio per il processamento di una lingua straniera e al maggiore ricorso ai sistemi di controllo del feedback articolatorio-uditivo in risposta a un input linguistico impegnativo di una lingua non familiare.

L'incremento della capacità percettiva con l'esposizione audiovisiva all'input, anziché unicamente sonora, si spiega col fatto che il gesto articolatorio visibile si configura come il punto nel quale i due livelli di multimodalità del linguaggio verbale (acustico-visiva e acustico-articulatoria) convergono e grazie al quale il processamento sonoro risulta facilitato. Il gesto articolatorio visibile si pone quindi come un elemento prezioso non solo nell'ambito del processamento della lingua materna, ma anche nell'ambito del processamento e della produzione di suoni linguistici di una seconda lingua.

Proprio a partire dal processo automatico di integrazione audiovisiva tra input sonoro e gesto articolatorio, infatti, è possibile reperire un indizio esplicito per la creazione delle mappe articolatorio-uditive e uditivo-articulatorie funzionali al più accurato processamento dell'input sonoro, che può a sua volta riflettersi in un miglioramento della capacità di produzione linguistica orale.

MULTIMODALITÀ

1.1 Il linguaggio umano: un sistema di sistemi

La comunicazione tra sapiens si fonda su sistemi diversi, facenti capo a strutture differenti che ne consentono la produzione e la comprensione.

Come illustrano Stephen Levinson e Judith Holler (Levinson & Holler, 2014), la comunicazione umana va vista nel suo complesso come un sistema composto da sistemi che si sono evoluti e affiancati nel tempo e che nel loro reciproco integrarsi e affiancarsi supportano la comunicazione. Il linguaggio verbale è solo uno di essi, evolutivamente il più giovane, presente oggi insieme a sistemi più antichi. In questa stratificazione di sistemi, la gestualità e la vista occupano un posto più arcaico del parlato.

La comparsa delle strutture anatomiche e dei circuiti neurali idonei a consentire la produzione orale nella sua forma più articolata è infatti riconducibile a un periodo compreso tra i 100.000 e i 50.000 anni fa (Lieberman P. & McCarthy, 2007), molto dopo la comparsa del sapiens sulla terra, collocabile intorno ai 200.000 anni fa.

È difficile immaginare che il linguaggio umano si sia evoluto, nelle sue prime fasi, in un contesto diverso da quello dell'interazione vis-à-vis. Sebbene nel mondo alfabetizzato e tecnologico, e ancor di più in quello in emergenza pandemica,

siamo sempre più abituati a concepire e a vivere la comunicazione tra sapiens attraverso canali tecnologici (video chiamata, e-mail o sms, telefono, ecc.), in origine, una comunicazione efficace doveva quasi certamente contare interamente sulla sua dimensione corporea. Ed è in questo contesto che la visione e la gestualità acquistano un ruolo fondamentale.

L'ipotesi di Levinson e Holler (2014) è che la comunicazione fosse inizialmente affidata ai canali visivo e gestuale e che nel corso del tempo a questi due sistemi sia andato affiancandosi, sovrapponendosi, ma mai escludendoli, il linguaggio verbale. Il risultato di questa integrazione reciproca è un sistema di comunicazione multimodale in cui canali diversi, quali per esempio la voce, l'espressione facciale, lo sguardo e i gesti contribuiscono a creare e a trasferire significato.

Il ruolo comunicativo del sistema visivo si può apprezzare già a partire dalle principali strutture che lo compongono: l'occhio umano si contraddistingue da quello di tutti gli altri primati per la sua morfologia, che lo caratterizza come un occhio che sembra essersi evoluto non per facilitare il suo mimetizzarsi bensì per renderlo facilmente individuabile (Kobayashi & Kohshima, 2001).

La parte visibile della sclera dell'occhio umano si differenzia infatti da quella dei primati non umani perché non è pigmentata, ma bianca e più estesa, mettendo così l'iride pigmentata in risalto. La combinazione di questa morfologia tipica, insieme con la forma fortemente allungata orizzontalmente dell'occhio, sarebbe una prova del fatto che l'occhio umano non si sia evoluto per facilitare il mimetizzarsi dello sguardo, funzionale al mettersi al riparo dall'avvistamento dei

predatori, ma al contrario, il significato adattivo di questa morfologia sarebbe invece proprio quello di facilitare lo sguardo reciproco, che assume infatti un ruolo estremamente importante nelle interazioni umane, molto meno in quelle tra gli altri primati (Bard et al., 2005), (Tomasello et al., 2007).

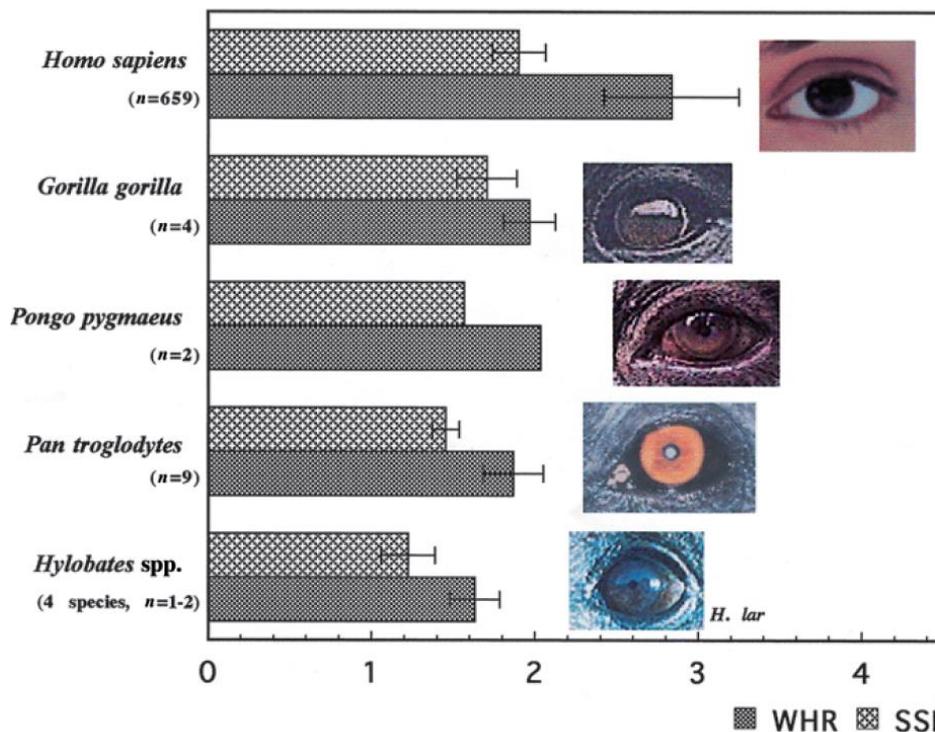


Figura 1. Variazione del rapporto tra larghezza e altezza (WHR) e indice di sclera visibile (SSI) nella superfamiglia degli Ominoidi (Kobayashi & Kohshima, 2001).

La capacità di stabilire il contatto visivo tra esseri umani è fondamentale per lo stabilirsi della relazione necessaria alla comunicazione (Gibson & Pick, 1963). E in effetti è stato dimostrato che fin dai primi giorni di vita (2-5 giorni) i neonati, nel guardare i volti, mostrano una sostanziale preferenza per i volti con sguardo diretto (“direct gaze”), rispetto ai volti il cui sguardo è rivolto verso destra o sinistra (“averted gaze”), nonché tempi di osservazione maggiori per i volti con

sguardo diretto (Farroni et al. 2002). Partendo da questi risultati, gli autori dello studio hanno ipotizzato, e verificato, un incremento dell'attività neurale per il processamento dei volti in bambini di 4 mesi quando essi vengono esposti a un volto con sguardo diretto anziché a un volto con sguardo rivolto lateralmente (Farroni et al. 2002).

Anche i bambini tra i 3 e i 6 mesi mostrano una preferenza per i volti con sguardo diretto, tanto che il sorriso che mantengono nelle fasi sperimentali in cui sono esposti a un volto adulto con sguardo diretto (periodi 1 e 3) tende a sparire nei periodi 2 e 4, in cui lo sguardo dell'adulto viene distolto da quello del bambino e rivolto lateralmente (Hains & Muir, 1996).

Quanto al sistema di comunicazione gestuale, la letteratura è ricchissima ed estremamente interessante, ma la sua ricostruzione dettagliata non rientra tra gli obiettivi di questo lavoro.

La gestualità interviene sia come elemento che ha valore per chi lo produce, a livello cognitivo e di elaborazione del pensiero sia a livello comunicativo per chi la osserva. David McNeill a metà degli anni '80 (McNeill, 1985) riaprì la strada a un dibattito, ancora vivo, sul rapporto che intercorre tra gesto e linguaggio verbale. Secondo McNeill (1985) i due sistemi, verbale e gestuale, sono tutt'altro che separati e procedono invece di pari passo condividendo anche i processi computazionali che li realizzano.

La visione di McNeill ha trovato i suoi oppositori, convinti per esempio che il gesto non proceda unicamente di pari passo al linguaggio ma anche in maniera

indipendente ad esso, per esempio supplendo al linguaggio parlato quando si incontra qualche difficoltà nell'espressione verbale (Feyereisen, 1987).

Come spesso accade, è il caso, soprattutto per l'analisi della gestualità e dei suoi rapporti con la cognizione e il linguaggio, di considerare un quadro d'analisi che integri i diversi aspetti in cui essa è coinvolta e di tenere in conto che l'approccio non può che essere variegato e flessibile, a seconda del contesto in cui si inserisce il gesto (Church & Goldin-Meadow, 2017). Oggi la gestualità è altresì al centro del dibattito riguardante l'evoluzione del linguaggio umano.

Secondo Corballis (2002), per esempio, sarebbe da rintracciarsi proprio in questo elemento l'origine del linguaggio così come lo conosciamo (Corballis, 2002).

Guardando allo sviluppo ontogenetico del linguaggio è largamente dimostrata la relazione tra gestualità e linguaggio verbale. I bambini indicano prima ancora di parlare, e la quantità dei significati trasmessi con i gesti nella fase preverbale è correlata alla successiva ampiezza del vocabolario. Allo stesso modo, il numero di combinazioni gesto-parlato a 18 mesi predice la complessità delle frasi usate a 42 mesi (Goldin-Meadow, 2000), (Goldin-Meadow et al., 2007).

Tra i 6 e i 9 mesi, in quella fase dello sviluppo linguistico definita lallazione (o, col termine inglese, "babbling"), in cui il bambino inizia ad articolare suoni sillabici più complessi rispetto ai suoni non linguistici tipici della fase precedente, detti "cooing", i suoni che vengono accompagnati ritmicamente dal movimento della mano hanno caratteristiche più simili a quelle dei suoni linguistici se paragonati ai suoni che il bambino emette senza l'associazione ritmica del movimento della mano (Ejiri & Masataka, 2001).

La comunicazione umana, dunque, si regge, si sviluppa e si compie a partire da più sistemi. Questo dato di multimodalità va sempre tenuto in conto nel linguaggio. Nelle lingue vocali, per esempio, i gesti co-verbali (quelli che accompagnano l'espressione verbale) integrano il significato espresso vocalmente e in modo simile, come illustra Fontana (2008), (2009), nelle lingue dei segni, le labializzazioni e i gesti orali possono intervenire ad integrare l'espressione segnata rendendola più efficace, a dimostrazione del fatto che la multimodalità del linguaggio, in ogni sua forma, rappresenta una risorsa estremamente preziosa sul piano comunicativo.

1.2 Meccanismi di acquisizione del linguaggio

Una delle tappe più attese dello sviluppo di un bambino è quella in cui compaiono le prime parole, quando i suoni emessi da un neonato iniziano pian piano a prendere la forma dei suoni a cui gli adulti della comunità linguistica di appartenenza attribuiscono valore linguistico. Col passare del tempo e col progressivo sviluppo delle strutture del tratto vocale che permettono una piena articolazione, i bambini riescono a produrre e combinare parole in maniera sempre più esperta.

Le prime tappe nella produzione linguistica seguono schemi riconoscibili in tutte le lingue (Guasti, 2007). Alle variazioni nel tipo di produzione vocale si affiancano cambiamenti nella struttura del sistema fonatorio, dalla cui maturazione dipende la possibilità di produrre suoni linguistici sempre più accurati.

Fino ai 4 mesi circa, i vocalizzi prodotti sono tendenzialmente suoni riflessi o associati ai bisogni primari. Intorno ai 4 o 5 mesi, la laringe comincia ad abbassarsi, la cavità orale inizia ad allungarsi e allargarsi e la varietà sonora cresce. I vocalizzi iniziano a riprodurre suoni vocalici e il bambino comincia a controllarne l'altezza e l'intensità (Guasti, 2007).

Successivamente, tra i 6 e gli 8 mesi, la capacità di coordinazione dei movimenti articolatori permette la produzione di suoni più vicini a quelli linguistici, ossia ripetizioni sillabiche semplici (come "mamama", "badaba", ecc.). Questa fase viene definita "babbling" o lallazione. La lallazione può essere canonica, quando si riflette nella produzione di una sequenza sillabica composta da una stessa sillaba ("mamama), o variabile, con la produzione di una sequenza di sillabe diverse tra loro ("badaba").

Questa è la fase a partire dalla quale la produzione di suoni inizia a somigliare di più al linguaggio, tanto che, tra gli 8 e i 10 mesi, si riscontrano differenze nel tipo di lallazione a seconda della lingua a cui il bambino è esposto, con una differenziazione nel tipo di suoni prodotti sulla base della lingua materna: i neonati tra gli 8 e i 10 mesi tendono a produrre con più frequenza le vocali e le consonanti presenti nella lingua a cui sono esposti (de Boysson-Bardies et al., 1989) e, anche a livello prosodico, il pattern dei suoni prodotti nella fase di lallazione tende a replicare quello della lingua di appartenenza (de Boysson-Bardies et al., 1984).

Le stesse tappe sono riscontrabili nei bambini sordi, che però a differenza dei bambini udenti attraversano la fase di lallazione manuale (Petitto & Marentette, 1991), (Meier & Willerman, 1995).

La fase preverbale è caratterizzata dal co-occorrere di gesti e vocalizzi, in cui entrambe le modalità attraversano fasi di sviluppo progressivo che conducono a un livello comunicativo via via più raffinato.

È stato riscontrato che tra i 6 e gli 11 mesi, nella fase della lallazione, vi è uno stretto legame tra i suoni e i gesti prodotti con le mani e le braccia: i vocalizzi accompagnati da gesti ritmici presentano sillabe più brevi e una durata minore dei passaggi tra frequenze formantiche se paragonati ai vocalizzi prodotti senza l'associazione di gesti ritmici (Ejiri & Masataka, 2001). Tali caratteristiche rendono i vocalizzi prodotti molto più simili ai suoni prodotti nel linguaggio verbale dei parlanti esperti, avvicinando i neonati a una produzione fonatoria di lallazione canonica sempre più linguisticamente connotata.

Il fatto che le caratteristiche sonore riscontrate nelle produzioni sillabiche realizzate in associazione a gesti ritmici permangano anche via via che la co-occorrenza del gesto ritmico va scomparendo segnala un probabile ruolo centrale di tali gesti nell'acquisizione di una capacità articolatoria più in linea con quella necessaria all'acquisizione del linguaggio verbale che richiede un movimento più rapido della glottide (la parte superiore della laringe, in corrispondenza delle corde vocali) e un gesto articolatorio più veloce (Ejiri & Masataka, 2001).

Tra i 9 e i 12 mesi, prima delle parole, i bambini iniziano a produrre gesti che assumono valore comunicativo. Tra i più rappresentativi e più comuni c'è il gesto

deittico di indicare, ossia il pointing (Bates, 1979), (Acredolo & Goodwyn, 1985). Il pointing è uno dei gesti più studiati della fase preverbale a causa della sua alta valenza comunicativa e dello stretto legame che si è riscontrato tra lo sviluppo del pointing e quello del linguaggio verbale al punto che si è rilevato che anomalie nella comprensione e nella produzione di pointing protodichiarativo (quello funzionale ad attirare l'attenzione altrui su ciò che viene indicato) sono collegate all'autismo (Baron-Cohen, 1989).

Nei bambini con sviluppo tipico i gesti non solo precedono l'espressione verbale, ma il loro andamento può anche predire il successivo andamento nello sviluppo linguistico: uno studio longitudinale condotto su bambini tra i 10 e i 24 mesi mostra che gli oggetti cui i bambini si riferiscono più spesso con i gesti nella fase preverbale o nella fase della comparsa delle prime parole, sono gli stessi che compaiono poi come elementi lessicali con lo sviluppo del linguaggio (Iverson & Goldin-Meadow, 2005). Il gesto resta similmente parte del processo di uno sviluppo linguistico più complesso anche quando il bambino è già in grado di articolare delle parole: i bambini che producono più gesti associati a una parola supplementare (per esempio, indicare la palla e dire "rossa") anziché una parola complementare (per esempio, indicare la palla e dire "palla") sono quelli in cui si manifesta prima la produzione di coppie di parole (Iverson & Goldin-Meadow, 2005). Inoltre, la quantità di combinazioni gesto-parlato prodotte a 18 mesi è correlata alla complessità delle frasi prodotte a 42 mesi (Goldin-Meadow, 2000), (Goldin-Meadow et al., 2007).

Lo sviluppo gestuale e quello linguistico (verbale) sono da considerarsi dunque come strettamente interconnessi. Essi sono talmente interconnessi che le

differenze di genere riscontrate nello sviluppo, per le quali i maschi tendono a produrre combinazioni di più parole più tardi rispetto alle femmine (Maccoby, 1966), (Kimura, 1999), si ritrovano anche nella produzione gestuale. Tra i 14 e i 34 mesi di vita, il numero e i tipi di gesti prodotti dai bambini maschi e dalle bambine femmine si eguagliano, si riscontra però una differenziazione nei tempi di comparsa della combinazione di gesto e parola supplementare (es. indicare il biscotto e dire “mangiare”): nei maschi, l’età media per la comparsa di questo tipo di combinazioni gesto + parlato è di 19,11 mesi, mentre nelle bambine si aggira intorno ai 16,36 mesi (Özçalışkan & Goldin-Meadow, 2010).

Gesti e parole non sono dunque in relazione tra loro semplicemente come strumento di compensazione l’uno dell’altra, ma formano insieme quel sistema comunicativo complesso, il cui sviluppo è un processo lungo che li vede intrecciarsi profondamente in un’interazione reciproca e dinamica (Campisi, 2018).

È verso i 12 mesi che la produzione linguistica comincia a vedere la comparsa delle prime parole, inizialmente con l’uso di protoparole: delle parole inventate ma con le quali il bambino si riferisce a un oggetto specifico, attribuendo a ciascuna protoparola un suo significato. Contemporaneamente si riscontra l’uso da parte del bambino di parole legate al contesto: esse somigliano a parole vere ma sono circoscritte all’uso in determinati contesti e sono generalmente comprese da chi ha più a che fare col bambino. Infine, le parole prodotte prendono la forma di suoni molto simili a quelli delle parole vere, sebbene non immediatamente pronunciate in maniera perfetta, e vengono usate in modo

accurato per riferirsi a un oggetto preciso in un contesto di interazione comunicativa (Guasti, 2007).

Il fatto che le prime parole compaiano, in maniera variabile e soggettiva, più o meno a partire dai 12 mesi, e che già tra le 24 e le 28 settimane di gestazione le strutture periferiche del feto lo predispongano al processamento di alcuni suoni linguistici (Werker & Gervain, 2013), implica che prima di raggiungere questo traguardo tanto atteso dai conspecifici adulti i bambini attraversino una lunga fase di sola percezione del linguaggio verbale e che durante questa fase processino gli elementi propedeutici alla successiva produzione linguistica.

Oltre a riconoscere e a preferire la voce della madre già prima dei 3 giorni di vita (DeCasper & Fifer, 1980), è stato dimostrato per esempio che i bambini, a partire dai 3 giorni di vita, riconoscono le frasi pronunciate dalla madre con frequenza giornaliera durante le ultime 6 settimane di gestazione (DeCasper & Spence, 1986). Nello studio, a partire dal settimo mese e mezzo, le madri dei bambini testati dopo la nascita, leggevano a voce alta, due volte al giorno e in una stanza silenziosa, la storia per bambini che era stata loro assegnata dagli sperimentatori. Dopo la nascita, i bambini del campione preferivano l'ascolto di un estratto appartenente alla storia letta dalla madre in gravidanza a quello di una storia mai ascoltata prima. Questa differenza nella preferenza per un estratto rispetto all'altro non era invece rilevata nei bambini del gruppo di controllo, ai quali nessuna delle due storie era stata letta in gravidanza. Inoltre i bambini di 2 giorni preferiscono ascoltare un parlante (sconosciuto) che si esprime nella loro lingua materna, quindi la lingua con cui hanno familiarità, anziché una a cui non sono mai stati esposti (Moon et al., 1993).

La lunga fase di solo processamento dei suoni linguistici è quindi importante e lascia dei segni. È grazie all'integrazione di meccanismi cognitivi differenti che, nella fase preverbale, si iniziano a processare e a discernere gli input linguistici in modo funzionale allo sviluppo delle capacità di produzione degli stessi. Come si vedrà, lo sviluppo ontogenico del linguaggio si può considerare come un processo in cui meccanismi generali percettivi e di apprendimento e meccanismi specializzati per il linguaggio si combinano e si integrano (Gervain & Mehler, 2010).

Va detto che il dibattito sui sistemi su cui si fonda l'acquisizione del linguaggio è estremamente vasto, e vede proposte e approcci teorici talvolta anche in contrasto tra loro. Una ricostruzione esaustiva di tale dibattito non rientra tra gli obiettivi di questo lavoro, pur riconoscendone il grande valore sia per la comprensione dei nodi problematici e teorici legati alla comprensione del processo specie-specifico del linguaggio umano, sia per prendere coscienza degli sforzi e dei progressi compiuti nella descrizione di tale processo, per nulla semplice da scomporre.

La linea qui seguita per fare luce sul processo di acquisizione linguistica è quella che fonda il processo di acquisizione sull'interazione tra meccanismi specifici per il linguaggio e meccanismi generali di apprendimento e di percezione che entrano in gioco anche per il linguaggio e che, in certi casi, il linguaggio stesso riflette nella propria struttura (Gervain & Mehler, 2010). Ne è un esempio la maggiore sensibilità dell'uomo alle ripetizioni e ai margini nelle sequenze sonore (Endress et al. 2009) e il fatto che le lingue collochino le ripetizioni e l'accento tonico

principalmente ai margini delle parole o comunque delle unità fonologiche (Low & McCarthy, 1983), (Hayes, 1995).

Un insieme di bias percettivi predispone i neonati a una preferenza per i suoni linguistici rispetto a quelli non linguistici (Vouloumanos & Werker, 2004), (Vouloumanos & Werker, 2007a), (Vouloumanos & Werker, 2007b).

Già dai primi giorni di vita, l'organizzazione corticale nel processamento dei suoni linguistici risulta a sua volta affine all'organizzazione dei cervelli adulti, con un'attivazione in risposta al linguaggio verbale nelle stesse regioni della corteccia perisilviana dell'emisfero sinistro (Dehaene-Lambertz et al. 2008) e attività sia nella circonvoluzione temporale superiore sia in quella angolare (Dehaene-Lambertz et al. 2013). L'organizzazione corticale ha infatti un assetto strutturale e gerarchico asimmetrico tale da rendere la corteccia uditiva sinistra specializzata nel processamento di suoni acustici rapidi, quali sono quelli del linguaggio verbale (Zatorre & Belin, 2001). Il processamento prosodico, che ha invece generalmente una lateralizzazione destra negli adulti, risulta ugualmente lateralizzato a destra nei neonati. La prosodia di una lingua è rappresentata da quegli elementi detti soprasegmentali, ossia quegli aspetti di una catena fonetica che non possono essere segmentati in porzioni di parole (e dunque trascritti) come la durata, l'intensità, il ritmo dei suoni linguistici e la frequenza fondamentale (f_0) della voce che è rappresentata dal numero di aperture e chiusure delle corde vocali in un secondo e a cui ci si riferisce nei termini di altezza della voce (Fabbro & Cargnelutti, 2018). L'aspetto prosodico è di particolare interesse, e lo si vedrà più avanti in questo paragrafo, perché è

proprio a partire dagli elementi contenuti nelle unità sonore del linguaggio che i bambini iniziano a riconoscere elementi strutturali della lingua a cui sono esposti.

L'organizzazione corticale sopra descritta agevola il riconoscimento dei suoni propri della lingua in cui il bambino è immerso, favorendone la progressiva specializzazione.

Al di là degli elementi fin qui esposti, che rendono l'essere umano più sensibile al linguaggio, ma che caratterizzano in maniera simile l'organizzazione cerebrale anche di altri primati (Gannon et al. 1998), bisogna però tenere conto di altri meccanismi che permettono ai neonati di farsi strada nella percezione, distinzione e categorizzazione dei suoni linguistici: tra questi, meccanismi generali di apprendimento e di individuazione di regole come l'estrazione statistica di informazioni significative e meccanismi di funzionamento della percezione e della memoria (Saffran et al., 1996), (Endress et al., 2009), (Werker & Gervain, 2013).

Il corpus linguistico a cui un bambino è esposto si presenta come un flusso sonoro pressoché continuo, tipico del linguaggio verbale. È in questo contesto che un neonato inizia a fare esperienza diretta della lingua (o delle lingue) di appartenenza ed è a partire da questo flusso sonoro che deve costruire la sua conoscenza e la sua competenza nel distinguere e segmentare le proprietà strutturali della lingua, necessarie alla progressione verso la fase di produzione. Avere una predisposizione innata per il processamento dei suoni linguistici, infatti, non implica la capacità automatica di categorizzare le unità strutturali della lingua a cui si è esposti. La capacità di riconoscere e categorizzare le unità

strutturali delle parole è però fondamentale allo sviluppo della capacità di produzione. Non può infatti esserci produzione linguistica senza che vi sia un riconoscimento preliminare delle unità strutturali della lingua che si vuole produrre.

Per spiegare il modo in cui un neonato riesce ad avviare il processo di segmentazione delle unità sonore strutturali della lingua in cui è immerso (una qualunque lingua del mondo), così da poter comprendere i confini e la forma delle strutture linguistiche di appartenenza, si sono avvicinate diverse possibili spiegazioni. La questione viene affrontata da vari autori rimandando a meccanismi di costruzione progressiva della conoscenza linguistica, in quello che Pinker (1984) chiama processo di “bootstrapping”. Tali meccanismi operano attraverso sistemi di apprendimento euristico che estrapolano le correlazioni tra i contenuti percettivi di una lingua e le sue proprietà morfologiche e sintattiche.

Il sostantivo “bootstrap” in inglese denota il tirante degli stivali, ossia le fibbie cucite ai lati o sul retro dello stivale per agevolarne la calzata. Il verbo “to bootstrap” si usa per connotare un’acquisizione personale ottenuta senza aiuti esterni, verso la quale si progredisce grazie alle proprie risorse personali che permettono passo dopo passo di ottenere risultati sempre maggiori.

Relativamente al bootstrapping linguistico esistono diverse prospettive, ripercorse qui molto brevemente, che individuano in elementi diversi gli indizi chiave su cui un bambino costruisce man mano le giuste correlazioni tra i suoni della lingua e le unità strutturali significative della stessa, in modo funzionale all’avvio di identificazioni categoriali sempre più ampie.

Pinker (1984), (1987) suggerisce che il bootstrapping parta dagli indizi di natura semantico-concettuale: a partire dalla comprensione di frasi semplici e dal significato di alcune parole, il bambino costruisce alberi sintattici da cui può man mano inferire le regole sintattiche della lingua di appartenenza, cosa che gli permette a sua volta di processare e comprendere frasi più complesse.

Il bootstrapping sintattico (Gleitman, 1990), (Gleitman & Landau, 1994) individua invece nell'inferenza sintattica il mattone iniziale per la categorizzazione linguistica e sintattica.

Entrambe queste visioni concettualizzano una costruzione categoriale che parte da una conoscenza linguistica preliminare già acquisita, per poter così procedere con un bootstrapping sempre più ad ampio raggio. Come notano Gervain e Mehler (2010), però, per capire come vengano acquisite queste conoscenze preliminari è utile volgere lo sguardo verso un bootstrapping incentrato sugli aspetti acustici e fonologici del linguaggio verbale. A partire dalle informazioni acustiche contenute nel linguaggio è infatti possibile estrarre informazioni morfosintattiche (Morgan & Demuth, 1996), (Nespor et al., 1996), (Ramus et al., 1999), (Nespor et al., 2008), (Mehler et al., 2004) e i neonati si mostrano già molto sensibili alla percezione di questi elementi.

Diversi studi contribuiscono con evidenze sperimentali a questa visione. Per esempio, già a 2 mesi i bambini sembrano avvalersi degli indizi prosodici del flusso continuo del verbale per organizzare e codificare alcune proprietà linguistiche, tanto che memorizzano maggiormente le proprietà fonetiche di

parole presentate in una frase rispetto a quelle di parole presentate in isolamento come liste di parole (Mandel et al. 1994).

Tra i 4 giorni e i 2 mesi i bambini sono in grado di distinguere la loro lingua madre da un'altra lingua sulla base delle caratteristiche ritmiche (Mehler et al., 1988); a 3 mesi l'attività corticale registrata rispetto all'esposizione a un cambio fonetico non particolarmente marcato (sillabe /ba/ e /ga/) segnala la capacità di discernere le due sillabe in un tempo pari a 400ms (Dehaene-Lambertz & Dehaene, 1994).

Sempre a partire dalle caratteristiche sonore delle parole, bambini di 3 giorni sono in grado di distinguere gli elementi di due categorie costituite da sostantivi, verbi, aggettivi, avverbi (nello studio "content words") da un lato e articoli preposizioni e ausiliari (nello studio "function words") dall'altro (Shi et al. 1999).

Esistono poi meccanismi percettivi e mnemonici generali che facilitano (e vincolano) la decodifica e la categorizzazione del linguaggio e che le lingue sembrano in qualche misura rispecchiare al loro interno. Tra questi, le ripetizioni e la salienza percettiva delle estremità delle parole: il sistema cognitivo umano mostra infatti una accentuata sensibilità verso questi elementi (Endress et al. 2009).

I neonati mostrano effettivamente una risposta neurale maggiore durante l'ascolto di parole con all'interno una ripetizione di sillabe (ABB) rispetto a parole con sequenze casuali di sillabe (ABC) (Mehler et al. 2008) e le ripetizioni tendono a essere collocate nella stragrande maggioranza delle lingue alle estremità (iniziali o finali) delle parole o comunque alle estremità di unità fonologiche (Low

& Mccarthy, 1983). Anche l'accento tonico tende a collocarsi sempre in relazione ai margini delle parole (Hayes, 1995).

A queste caratteristiche delle lingue corrisponde la capacità dei neonati di identificare la collocazione delle sillabe in lunghe sequenze sillabiche (nello studio, sequenze da 6 sillabe) solo in relazione al cambiamento occorso nel posizionamento delle sillabe poste alle estremità della sequenza (Ferry et al., 2016). Dopo una fase di familiarizzazione con le sequenze multisillabiche presentate, l'aumento di concentrazione di emoglobina ossigenata segnala la sensibilità rispetto al cambiamento dell'ordine sillabico solo quando questo riguarda le sillabe alle estremità della sequenza, non quando riguarda le sillabe centrali.

La percezione dell'inversione nella posizione delle due sillabe centrali viene percepita però nel caso in cui tra le due sillabe intercorra una pausa di 25ms, elemento che avvalorava l'ipotesi che indizi anche deboli nel flusso linguistico possano rappresentare chiavi di accesso alla sua segmentazione (Ferry et al., 2016).

Quanto ai meccanismi generali di acquisizione, uno di quelli maggiormente presi in considerazione dalle posizioni empiriste sull'acquisizione del linguaggio è quello dell'apprendimento per estrazione statistica, che permette di ricavare informazioni utili alla codifica e alla categorizzazione delle unità linguistiche. Tale sistema non è considerato un elemento specie-specifico votato al linguaggio umano, ma fa parte invece di quei meccanismi generali che l'uomo condivide con altri primati, con evidenze tra i primati non umani anche nel caso della

categorizzazione linguistica (Hauser et al. 2001). Rispetto al ruolo dell'apprendimento statistico nell'acquisizione linguistica sono state prodotte numerose evidenze. I neonati, così come i bambini di 8 mesi, sono in grado di estrarre le proprietà statistiche del flusso continuo del parlato in modo tale da riuscire a individuare i confini delle parole (Saffran et al. 1996), (Teinonen et al., 2009).

È possibile che inizialmente i neonati facciano grande affidamento sui meccanismi generali di estrazione di regole, in particolare quelli basati sulle informazioni statistiche per poi riservare via via maggiore attenzione a elementi sempre più linguisticamente connotati (Werker & Gervain, 2013).

Bisogna inoltre sempre tenere a mente che, insieme con i meccanismi che rendono possibile il processamento, il riconoscimento, la categorizzazione e la produzione dei suoni linguistici, l'atto linguistico è legato anche, in maniera imprescindibile, alla struttura corporea che rende possibile questa performance così variegata, ossia il tratto vocale sopralaringeo (Pennisi & Falzone, 2016).

Sulla base di quanto finora descritto, quindi, possiamo affermare che anche nel caso dei meccanismi di acquisizione del linguaggio non si può prescindere da una visione multimodale, all'interno della quale tutti i meccanismi fin qui introdotti sono da considerarsi in un'ottica sinergica di integrazione e interazione reciproca. Questo processo di acquisizione si realizza grazie a una serie di competenze di base che coinvolgono più sistemi percettivi e più modalità di generalizzazione inferenziale.

Come vedremo nel prossimo paragrafo, la multimodalità nell'acquisizione del linguaggio trova fondamento proprio nei sistemi di percezione del linguaggio verbale durante l'infanzia.

1.3 La percezione del linguaggio verbale durante l'infanzia: un processo multisensoriale

Come si è visto nel primo paragrafo, la natura multimodale del linguaggio è un prezioso connotato dei sistemi linguistici che contribuisce a renderli più funzionali alla comunicazione. Nelle lingue vocali, pur rappresentando solo uno dei sistemi di segnalazione per l'espressione del pensiero (Fitch, 2010), il linguaggio verbale assume spesso una sorta di primato sugli altri sistemi di segnalazione da cui deriva la considerazione del valore dominante del senso dell'udito per il processamento linguistico. È sbagliato però considerare il linguaggio verbale come una modalità comunicativa unisensoriale. Il processamento del linguaggio verbale si affida infatti anche al canale visivo, che conferisce allo stimolo verbale una natura multisensoriale. L'uomo ha la capacità di integrare in maniera coerente informazioni provenienti da sensi diversi ma relativi allo stesso evento (Stein & Meredith, 1993) e studiare il linguaggio verbale quale fenomeno multisensoriale, anziché unicamente uditivo, permette di comprenderne aspetti talvolta trascurati che hanno tuttavia una loro rilevanza nel contesto dell'appropriazione linguistica.

Nella fase preverbale i bambini rispondono in maniera preferenziale a eventi multisensoriali riconducibili a una sorgente unica, come per esempio la vista della caduta di un giocattolo a terra e il suono a essa correlato (Murray et al., 2016).

La percezione multisensoriale è generalmente più ricca rispetto a quella unimodale (Ernst & Banks, 2002), facilitando così la possibilità di apprendere dall'input (Bremner et al. 2012), (Hatwell, 2003).

In quest'ottica, il linguaggio verbale si presenta come candidato ideale, in quanto stimolo preferenziale e saliente: l'input sonoro del parlante è infatti indissolubilmente legato al gesto articolatorio che ne permette la produzione e che è, in condizioni normali, congruente all'input linguistico sonoro, rispettando dunque la condizione di congruenza spazio-temporale che garantisce l'integrazione dell'input verbale sonoro con quello visivo (Jack & Thurlow, 1973), (Dodd & Campbell, 1987), (Wallace et al., 2004).

Il gesto articolatorio è l'atto motorio che coinvolge il tratto vocale e i muscoli facciali e che consente la produzione linguistica. La parte visibile del gesto articolatorio è rappresentata dal movimento della bocca di un parlante e costituisce l'input visivo che è strettamente connesso all'input sonoro del linguaggio verbale.

Il volto del parlante costituisce un contesto percettivo che arricchisce l'input linguistico e già per i neonati l'esperienza di percezione linguistica può configurarsi come un processo visivo, oltre che uditivo: tra le 9 e le 12 settimane di vita, i neonati hanno capacità di associare il gesto articolatorio (di volti femminili e maschili) all'input sonoro vocalico relativo (Bristow et al., 2009).

Il volto del parlante contiene indizi preziosi per i piccoli umani, che a 6 mesi replicano con più frequenza i suoni vocalici, emessi da un volto presentato in posizione dritta (anziché capovolta) e con sguardo diretto (anziché rivolto lateralmente). L'aumento della riproduzione del suono è connesso anche ai tempi di osservazione della bocca del parlante. Maggiore è il tempo in cui i bambini guardano la bocca che articola i suoni linguistici, maggiore è la riproduzione dei suoni (Imafuku et al., 2019).

Anche negli adulti è presente una stretta integrazione cross-modale per il processamento di suoni linguistici e gesto articolatorio, con l'utilizzo delle informazioni fonetiche rilevanti offerte dall'input visivo (Brancazio & Miller, 2005).

L'atto verbale, dunque, contiene al suo interno una forte connotazione corporea legata al movimento articolatorio, che lo rende più ricco e difficilmente concepibile come atto linguistico unidimensionale, anche quando preso in considerazione singolarmente come sistema di segnalazione isolato dagli altri.

Nei successivi due capitoli si approfondirà il ruolo dell'input visivo del linguaggio verbale, nello specifico il gesto articolatorio (o movimento articolatorio della bocca), nelle diverse fasi dello sviluppo.

IL CORPO COME CONTESTO PERCETTIVO: IL GESTO ARTICOLATORIO DURANTE L'INFANZIA

2.1 L'integrazione audiovisiva nei neonati

Il ruolo comunicativo del sistema visivo durante l'infanzia non si limita a quello di canale relazionale per il contatto visivo: esso ha un ruolo importante anche nella percezione e nel processamento dei suoni del linguaggio verbale.

I neonati, molto più degli adulti, possono contare su un'interazione con i conspecifici che ha luogo prevalentemente in un contesto di interazione faccia a faccia, diretto e non mediato da supporti tecnologici. Sono quindi immersi pienamente nella dimensione multimodale della comunicazione ed è a partire da questa dimensione che iniziano a fare esperienza del linguaggio umano nel suo complesso.

Tale multimodalità va considerata sia nella produzione sia nel processamento del linguaggio ed è visibile già guardando alle prime fasi percettive nei neonati.

La produzione di parole è inscindibile dal movimento della bocca di chi le pronuncia e fin dai primi mesi di vita i bambini integrano l'input visivo offerto dal movimento articolatorio con quello sonoro a esso associato (Dodd, 1979), (Kuhl

& Meltzoff, 1982), (Rosenblum et al., 1997), (Patterson & Werker, 2003), (Denis Burnham & Dodd, 2004).

Uno studio del 1979 (Dodd, 1979) condotto su bambini tra le 10 e le 16 settimane (2.5 - 4 mesi), in cui gli sperimentatori, una volta attirata l'attenzione del bambino durante una breve fase di familiarizzazione, recitavano delle filastrocche e in cui il sonoro relativo alle parole pronunciate veniva presentato talvolta in sincronia col movimento della bocca, talvolta in asincronia rispetto al relativo gesto articolatorio, rilevò livelli di disattenzione maggiori (minor tempo di osservazione del volto del parlante) per i volti che presentavano la condizione di asincronia tra suono e gesto articolatorio, portando gli autori a concludere che i neonati a questa età hanno già la capacità di riconoscere la relazione corretta tra gli stimoli verbali sonori e articolatori prodotti da un parlante preferendola a una associazione incongruente dei due input (Dodd, 1979).

Uno studio successivo (Kuhl & Meltzoff, 1982), condotto su bambini tra le 18 e le 20 settimane (4.5 - 5 mesi), registra la capacità dei lattanti di percepire la corrispondenza tra le informazioni sonore e visive del linguaggio verbale, avvalorando la tesi che la percezione del linguaggio verbale non sia per nulla un fenomeno confinato unicamente alla percezione degli aspetti uditivi.

Non si tratta più di suoni sincroni e asincroni rispetto al gesto articolatorio, ma di due suoni vocalici diversi (/a/, /i/) presentati insieme a due volti affiancati, di cui uno articola il suono /a/ e l'altro il suono /i/.

B		Familiarization		Test	
Visual stimuli	 Face 1	 Face 2	Midline gaze	 Both faces	
Auditory stimuli		/a/.../a/.../a/.../a/	
Time	10 seconds	10 seconds		2 minutes	

Figura 2. Procedura sperimentale (Kuhl & Meltzoff, 1982).

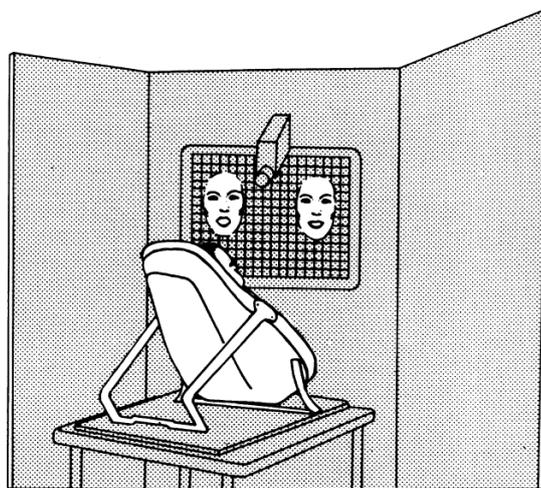


Figura 3. Setting sperimentale con la disposizione del neonato sul seggiolino posto in un box con tre pareti a 46 cm dal display con i 2 volti (Kuhl & Meltzoff, 1982).

Gli studi citati utilizzavano solo volti e voci femminili, cui nei primi mesi di vita (e durante la vita intrauterina per quanto riguarda la voce) il bambino è più esposto.

Alla luce di ciò, Michelle Patterson e Janet Werker (Patterson & Werker, 1999) condussero un nuovo studio per verificare se la percezione bimodale (visiva e uditiva) del linguaggio verbale nei primi mesi di vita fosse generalizzabile anche usando nei test volti maschili e introducendo volti più naturali. A differenza degli studi precedenti, in cui il volto del parlante era rappresentato in isolamento, con

i soli contorni del viso, in questo studio i parlanti sono rappresentati con tutte le parti del corpo dalle spalle in su.

Anche in questo caso, i due input sonori utilizzati sono quelli delle vocali /a/ e /i/, associati al gesto articolatorio corrispondente o a quello della vocale alternativa.

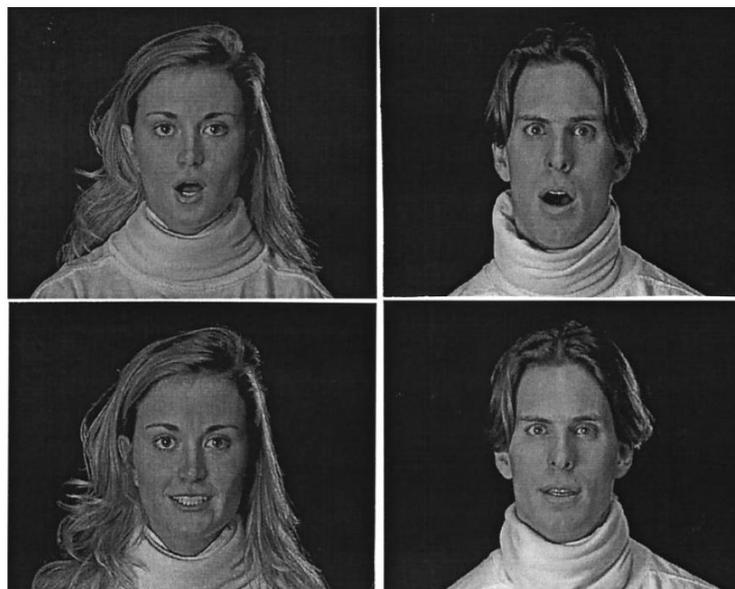


Figura 4. Volti usati per l'articolazione dei suoni /a/ e /i/ (Patterson & Werker, 1999).

I test condotti nell'ambito di questo studio confermano i risultati sulla preferenza per i volti in cui il movimento articolatorio corrisponde al suono prodotto rispetto ai volti in cui il suono vocalico prodotto è discordante rispetto al gesto articolatorio, con tempi individuali di osservazione più lunghi per i primi e una media del 64.7% del tempo totale di osservazione a favore dei volti in cui la corrispondenza movimento articolatorio e suono è rispettata, senza alcuna differenza significativa tra volti maschili e femminili, dato da cui le autrici

ricavano la conclusione che elementi quali la frequenza fondamentale e la struttura formantica dei suoni non influiscono sulla capacità di associare le informazioni visive e acustiche emesse dal parlante.

Tra i risultati di questo studio, condotto su bambini di 4 mesi e mezzo, le autrici riportano anche un dato relativo alla riproduzione dei suoni vocalici ascoltati dai bambini: i bambini esposti a un volto in cui il gesto articolatorio e il suono emesso concordavano riproducevano più a lungo i suoni ascoltati (Patterson & Werker, 1999).

Le stesse autrici replicarono i risultati ottenuti in uno studio successivo (Patterson & Werker, 2003), condotto questa volta con bambini di 2 mesi di età, per valutare se la preferenza per i volti con gesto articolatorio e input sonoro congruenti fosse riscontrabile anche a uno stadio ancora più iniziale della vita.

La media del totale di tempo di osservazione dei volti presentati segnala che il 63.95% del tempo era stato impiegato nel guardare i volti in cui suono emesso e gesto articolatorio corrispondevano.

Durante la presentazione dei due volti (volto con suono corrispondente al gesto articolatorio e volto con suono non corrispondente) i bambini guardavano per tempi notevolmente più lunghi il volto con movimento della bocca e suono corrispondenti, confermando così i risultati precedenti ed estendendoli a una fascia d'età più bassa, quella dei bambini di 2 mesi (Patterson & Werker, 2003).

2.2 Il ruolo del gesto articolatorio nel processamento dei suoni linguistici

La capacità dei neonati di processare suoni linguistici partendo dall'integrazione audiovisiva dell'input ha risvolti importanti sulla capacità di distinguere i diversi fonemi e quindi sul ruolo del gesto articolatorio del parlante come risorsa per la categorizzazione delle componenti sonore strutturali del linguaggio.

Tra i 5 e i 6 mesi i bambini hanno la capacità non solo di riconoscere la congruenza e la corrispondenza tra lo stimolo linguistico sonoro e quello visivo (gesto articolatorio), ma anche di distinguere fonemi molto simili a seguito dell'esposizione audiovisiva a essi, fallendo invece nella distinzione degli stessi fonemi quando il gesto articolatorio non contribuisce alla loro differenziazione (Teinonen et al., 2008).

Nello studio, 48 bambini tra i 5 e i 6 mesi (età media 6 mesi e 11 giorni) venivano esposti a fonemi compresi in un continuum di 8 fonemi tra /ba/ e /da/, due suoni difficili da distinguere perché acusticamente vicini. Nello studio i bambini erano divisi in due gruppi.

A uno dei gruppi (quello a categoria doppia) veniva presentato il gesto articolatorio per /ba/ o /da/ e uno dei suoni del continuum relativo a /ba/ o /da/ congruente. Nella fattispecie, un movimento articolatorio /ba/ era associato ai fonemi /ba1/ e /ba4/ e al movimento articolatorio /da/ erano associati i fonemi /da5/ e /da8/.

Al secondo gruppo (a categoria unica) venivano invece sottoposti entrambi i suoni del continuum in associazione a un unico gesto articolatorio /ba/ o /da/.

Dopo la fase di familiarizzazione veniva testata la capacità di distinguere i suoni del continuum (nello specifico /ba3/ e /da6/), quando presentati solo acusticamente.

I risultati mostrano l'incapacità dei bambini del gruppo a categoria unica di distinguere i due fonemi.

I risultati dei bambini del gruppo a categoria doppia mostrano invece la capacità di discernimento dei due fonemi presentati in modalità solo sonora nella fase di test. Questo anche nonostante i due suoni della fase di test (/ba3/ e /da6/) siano quelli più difficili da distinguere in maniera netta come appartenenti all'una o all'altra categoria fonetica (/ba/ o /da/).

Alla luce del ruolo chiave della componente visiva nella distinzione dei fonemi, uno studio successivo condotto su bambini di 18 mesi ha esteso l'indagine alla verifica della possibilità che i soli indizi visivi veicolino l'apprendimento di parole nuove. Altro scopo dello studio era verificare se parole apprese in modalità solo acustica fossero poi comprese correttamente se presentate nella sola modalità visiva (Havy et al., 2017).

Le parole da imparare erano 4 coppie di pseudoparole (inventate riprendendo i suoni dell'inglese) presentate in associazione a immagini corrispondenti all'oggetto indicato dalla parola. La presentazione delle nuove parole, seguite poco dopo dall'immagine dell'oggetto al centro del monitor, avveniva in modalità visiva (solo gesto articolatorio) o in modalità solo acustica.

Nella fase di test, gli oggetti presentati nella fase di apprendimento erano mostrati in coppia e il bambino doveva capire a quale oggetto fosse riferita la parola pronunciata acusticamente o visivamente. Questa fase prevedeva che i bambini fossero testati nella stessa modalità a cui erano sottoposti durante la fase di training (acustica o visiva), o in modalità diversa da quella del training (acustica per i bambini allenati con presentazione visiva e visiva per i bambini allenati in modalità acustica).

I risultati mostrano che i bambini allenati in modalità unicamente acustica riescono a imparare l'associazione parola-oggetto partendo dall'informazione sonora sia quando testati col solo stimolo visivo sia col solo stimolo acustico. I bambini allenati in modalità visiva invece falliscono nel compito in entrambe le modalità di test, visiva e sonora. A 18 mesi quindi il solo input articolatorio non è sufficiente ad apprendere parole nuove.

È interessante, tuttavia, notare che i bambini allenati in modalità acustica hanno la capacità di crearsi, a partire dal solo suono linguistico, la rappresentazione del gesto articolatorio a esso associato pur non avendolo mai visto prima.

Nella fase di test in modalità visiva, infatti, i bambini allenati con il solo input sonoro riuscivano ad associare con successo la parola presentata all'oggetto corrispondente.

Questi risultati mostrano che il gesto articolatorio può facilitare, anche in modo determinante, l'accuratezza del processamento linguistico nei neonati, ma anche che a 18 mesi i bambini sono in grado di estrapolare la rappresentazione del gesto articolatorio associato a una parola nuova che hanno solamente sentito.

2.3 Bilingui vs monolingui: riconoscimento visivo del linguaggio verbale

Uno studio del 2007 estende la rilevanza del gesto articolatorio alla capacità di distinguere lingue differenti facendo ricorso alla sola osservazione del movimento articolatorio, senza che esso sia accompagnato dal relativo input sonoro (Weikum et al., 2007).

Per fare ciò Weikum e colleghi (2007) usarono video muti di bilingui (inglese/francese) intenti a produrre frasi in ciascuna delle due lingue e misurarono i tempi di osservazione per le frasi in francese dei bambini monolingui (inglese) di 4 mesi facenti parte del campione, concludendo che a questa età i bambini sono in grado di differenziare, facendo ricorso ai soli stimoli visivi (gesto articolatorio), la loro lingua madre da una lingua che non conoscono.

Nello stesso studio gli autori paragonarono i risultati ottenuti dai bambini monolingui (inglese) e bilingui (inglese/francese) di 6 e 8 mesi. Per bilingui si intendono qui i bambini esposti fin dalla nascita a due lingue differenti.

Ne risulta che fino a 6 mesi entrambi i gruppi mostrano tempi di osservazione prolungati per le espressioni in lingua francese (quella non familiare per i monolingui), a dimostrazione del fatto che entrambi i gruppi differenziano visivamente le due lingue. Ma, a 8 mesi, questo incremento dell'attenzione rispetto al cambiamento di lingua (a indicare la percezione del cambiamento) si registra solo nei bilingui (Weikum et al., 2007).

Utilizzando le stesse procedure di Weikum e colleghi (Weikum et al., 2007), uno studio successivo indaga la possibilità che la capacità dei bilingui di 8 mesi di differenziare visivamente le due lingue della loro coppia linguistica sia una competenza generalizzabile alla capacità di distinguere visivamente anche lingue diverse da quelle a cui i suddetti bambini sono esposti normalmente (Sebastián-Gallés et al., 2012).

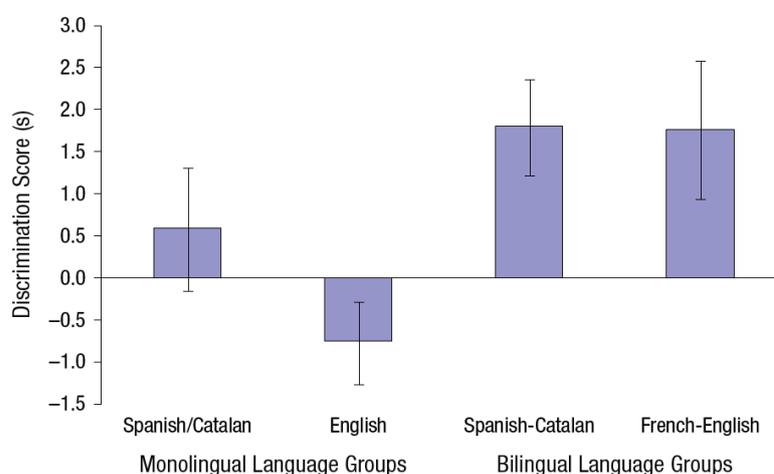


Figura 5. Punteggio medio di riconoscimento visivo delle lingue proposte (Sebastián-Gallés et al., 2012).

In questo caso a essere testati erano bambini bilingui spagnolo-catalano e bambini monolingui spagnoli, o catalani tra i 4 e gli 8 mesi.

Gli individui dei due gruppi venivano esposti a videoregistrazioni mute di parlanti bilingui inglese-francese.

I risultati ottenuti, poi comparati con quelli ottenuti da Weikum e colleghi (Weikum et al., 2007), mostrano che la capacità dei bilingui spagnolo-catalano di

distinguere visivamente lingue diverse è generalizzabile alle lingue che non appartengono alla loro coppia linguistica. Al contrario, i monolingui di 8 mesi non mostrano la capacità di distinguere le differenti lingue (misurata sulla base dei tempi di osservazione del volto parlante).

A 8 mesi, quindi, i monolingui perdono la capacità di distinguere visivamente la loro lingua da un'altra, mentre la capacità dei bilingui a 8 mesi di distinguere lingue diverse non si limita alla distinzione tra le lingue a cui sono esposti normalmente, ma si estende a lingue per loro nuove.

Il fatto che i bilingui di 8 mesi siano in grado di distinguere visivamente non solo le lingue conosciute ma anche lingue a loro estranee (a differenza dei monolingui della stessa età) è stato interpretato dagli autori dello studio (Sebastián-Gallés et al., 2012) come indicativo del fatto che, nei bilingui, la capacità di discernere visivamente lingue diverse non derivi da un processo di specializzazione per le due lingue a cui sono esposti e non si limiti a queste, lasciando pensare che non si tratti del risultato di un restringimento percettivo ("perceptual narrowing") dovuto all'esperienza fatta delle due lingue, ma piuttosto di un incremento dell'attenzione percettiva ("perceptual attentiveness").

Tuttavia, nei monolingui adulti, si registra la capacità di distinguere visivamente, quindi sulla base del solo gesto articolatorio, la propria lingua da un'altra, anche se con un livello di precisione minore rispetto ai bilingui della coppia linguistica che viene presentata visivamente. Questa abilità non è più rilevata però nei monolingui quando tra le due lingue presentate visivamente non c'è la loro lingua materna (Soto-Faraco et al., 2007).

Ciò significa che il restringimento percettivo osservato nei monolingui di 8 mesi rispetto alla capacità di distinguere visivamente la propria lingua da un'altra (Weikum et al., 2007), (Sebastián-Gallés et al., 2012) viene superato in età adulta probabilmente in relazione alla maggiore esperienza con il processamento audiovisivo della propria lingua materna e quindi a una maggiore conoscenza dei suoi tratti articolatori distintivi.

2.4 Attenzione selettiva alla regione della bocca

Il vantaggio dei bilingui nel compito di distinzione visiva di lingue differenti indica probabilmente che essi sono più abili o abituati a rivolgere la loro attenzione agli indizi articolatori del linguaggio verbale.

Questo potrebbe forse essere un indicatore del fatto che l'esperienza della doppia esposizione linguistica moduli l'attenzione selettiva alla regione della bocca del parlante in maniera diversa rispetto all'esperienza monolingue.

Uno studio longitudinale condotto su bambini di 6 mesi registra la correlazione tra tempi di osservazione della bocca del parlante (nello studio la madre del bambino) e sviluppo del linguaggio espressivo (usato per comunicare pensieri, idee e bisogni) a 18 mesi. A tempi di osservazione maggiori della regione della bocca del parlante a 6 mesi corrisponde uno sviluppo maggiore del linguaggio espressivo del bambino a 18 mesi (Young et al., 2009).

Uno studio successivo, che coinvolge bambini di 12 mesi registra la correlazione tra maggiori tempi di osservazione della regione della bocca di un parlante e ampiezza del vocabolario a 18 e 24 mesi (Tenenbaum et al., 2015)

Lewkowicz e Hansen-Tift (2012) hanno condotto uno studio proprio per rilevare le eventuali differenze nella direzione dello sguardo dei bambini nei primi 12 mesi di vita mentre guardano e ascoltano una donna recitare un monologo nella loro lingua madre (l'inglese) e in una lingua a loro estranea (lo spagnolo). I risultati hanno fornito uno schema comportamentale pressoché fisso e quindi generalizzabile.

Si sono infatti riscontrati dei passaggi comuni nell'attenzione selettiva dello sguardo dei bambini coinvolti, permettendo ai due studiosi di stabilire delle tappe nella preferenza occhi/bocca del parlante nel primo anno di vita.

Fino ai 4 mesi l'attenzione è rivolta prevalentemente agli occhi del parlante, successivamente, tra i 4 e gli 8 mesi, l'attenzione selettiva si sposta alla regione della bocca, indipendentemente dalla lingua usata dal parlante.

Raggiunti invece i 12 mesi, l'attenzione selettiva torna a rivolgersi agli occhi quando il parlante utilizza la lingua madre del bambino, ma non quando utilizza una lingua non familiare (nella fattispecie, lo spagnolo), momento in cui l'attenzione si sposta alla regione della bocca.

Risultati simili sui monolingui sono stati ottenuti anche da Pons e colleghi (Pons et al. 2015) con bambini monolingui catalani e spagnoli, e quindi abituati a caratteristiche ritmiche della lingua diverse dall'inglese.

Questo studio estende l'osservazione a un gruppo di 63 bambini bilingui (spagnolo-catalano). Nell'analizzare le differenze nell'attenzione selettiva per le diverse aree del volto di monolingui e bilingui nel primo anno di vita, gli autori della ricerca rilevano una preferenza generale dei bilingui per la regione della bocca.

In particolare, a 4 mesi i monolingui guardano maggiormente gli occhi, mentre i bilingui della stessa età rivolgono l'attenzione in misura uguale agli occhi e alla bocca del parlante, a indicare, per gli autori, uno spostamento precoce dell'attenzione selettiva dagli uni all'altra. A 8 mesi, infatti, entrambi i gruppi guardano maggiormente la bocca anziché gli occhi.

A 12 mesi invece sono i monolingui a guardare in egual misura gli occhi e la bocca quando esposti a un volto che parla la loro lingua materna: al contrario guardano più a lungo la bocca se il parlante utilizza una lingua a loro estranea.

I bilingui di 12 mesi invece rivolgono l'attenzione maggiormente e più a lungo alla bocca sia nel caso di una lingua familiare sia nel caso di una non familiare.

I risultati ottenuti supportano l'ipotesi che i bilingui siano maggiormente orientati verso la regione della bocca probabilmente per trarre un vantaggio ulteriore dagli indizi visivi messi a disposizione dal gesto articolatorio del parlante (Pons et al., 2015).

IL CORPO COME CONTESTO PERCETTIVO: IL GESTO ARTICOLATORIO DOPO L'INFANZIA

3.1 Gesto articolatorio e processamento del linguaggio verbale

La multimodalità nella percezione del linguaggio verbale e, nel caso in esame, la bimodalità data dall'integrazione degli aspetti sonori (voce) con gli aspetti visivi (gesto articolatorio) è una caratteristica che può influenzare il modo in cui i suoni vengono percepiti.

Partendo da questa constatazione, nata dall'osservazione del fatto che adulti cui veniva mostrato il video di una donna sul cui movimento della bocca per la sillaba [ga] era stato doppiato il sonoro della sillaba [ba] riferivano di sentire il suono [da], McGurk e MacDonald (1976) portarono avanti uno studio per approfondire e indagare l'effetto riscontrato, che prese poi proprio il nome di "McGurk effect".

Gli sperimentatori divisero le persone dello studio in tre gruppi così composti: un gruppo di 21 bambini dai 3 ai 4 anni; un gruppo di età compresa tra i 7 e gli 8 anni composto da 28 individui e un gruppo di 54 adulti tra i 18 e i 40 anni, con una

prevalenza di individui di sesso maschile. Nei due gruppi di bambini il rapporto maschio-femmina era invece più o meno equilibrato.

Il materiale preparato per lo studio consisteva nel filmato del volto di una donna mentre ripeteva le sillabe appaiate ba-ba, ga-ga, pa-pa, ka-ka. Da questo video fu creato un secondo video, in cui al movimento delle labbra della donna intenta a pronunciare ogni coppia di sillabe veniva associata la seguente combinazione di voce e gesto articolatorio: ba-voce/ga-labbra, ga-voce/ba-labbra, pa-voce/ka-labbra, ka-voce/pa-labbra. In ciascun video l'associazione sillaba-gesto articolatorio veniva replicata tre volte.

I soggetti dello studio venivano testati sulla comprensione dei suoni sopracitati in due condizioni: solo uditiva e audiovisiva. Tutti gli individui furono testati individualmente in entrambe le condizioni e su ciascuna delle quattro opzioni sonore. Ne risultò che nella condizione esclusivamente uditiva le percentuali per i suoni interpretati correttamente era molto alta: il 91% nel gruppo 3-4 anni, il 97% nel gruppo 7-8 anni e il 99% nel gruppo di adulti tra 18 e 40 anni.

Gli errori si rivelarono invece in percentuali significative per i test condotti nella condizione audiovisiva: 59% di errori nel gruppo 3-4 anni, 97% per il gruppo 7-8 anni e 99% di errori per il gruppo di adulti.

Gli autori suddivisero le risposte in due categorie "fused" (quando l'informazione ottenuta dalle due modalità veniva trasformata in un elemento nuovo, non presente nelle due modalità) e "combination" (quando la risposta era un composto di elementi relativamente immutati presi da ciascuna modalità). Le

risposte che non rientravano in nessuna di queste categorie furono etichettate con “other”.

I soggetti adulti diedero il 98% di risposte “fused” per la coppia audiovisiva ba-voce/ga-labbra e il 59% di risposte “combination” per la coppia a essa complementare (ga-voce/ba-labbra). Le percentuali si abbassano leggermente per la coppia pa-labbra/ka-voce, con l’81% di risposte “fused” da parte degli adulti e il 44% di risposte “combination” per la coppia complementare (ka-labbra/pa-voce).

Un dato interessante riguarda la modalità dominante nelle risposte in cui uno dei due canali percettivi (uditivo o visivo) influenza maggiormente la percezione del suono: in questi casi, il senso dominante nei bambini è l’udito, mentre le risposte degli adulti mostrano una dominanza della vista.

Fare click sull’immagine per avviare il video



Video 1. Esempio di effetto McGurk. Nel video, lo psicologo Lawrence D. Rosenblum, studioso di percezione e integrazione multimodale.

3.2 Il ruolo del gesto articolatorio in età adulta

Il legame (“binding”) dello stimolo cross-modale uditivo e visivo nella percezione del linguaggio verbale aumenta conseguentemente allo sviluppo (Ross et al., 2011) e per effetto dell’esperienza linguistica (Mercure et al., 2019), tanto che i bambini sono meno inclini a cadere nell’illusione legata all’effetto McGurk e mostrano comunque una maggiore influenza dell’input sonoro su quello visivo quando l’input audiovisivo presentato è incongruente (MacDonald & McGurk, 1976).

Negli adulti, relativamente al processamento audiovisivo del linguaggio, l’input visivo, che nella produzione verbale precede di qualche millisecondo l’input sonoro, sembrerebbe agire come acceleratore per il processamento acustico, producendo potenziali evocati acustici minori ma più veloci rispetto all’input linguistico presentato solo acusticamente. Questa facilitazione temporale diminuisce al diminuire della salienza dell’input visivo rispetto a quello sonoro, tanto da non essere registrata nei casi di input audiovisivo non coerente, come nel caso dell’illusione di McGurk (Van Wassenhove et al. 2005).

L’input visivo crea quindi una sorta di contesto percettivo nel quale collocare e processare l’input sonoro. L’effetto di facilitazione percettiva data dal gesto articolatorio non è però esclusivamente di ordine temporale.

In età adulta l’informazione data dal gesto articolatorio diventa rilevante e proficua anche a livello lessicale nel contesto dell’apprendimento di parole nuove. A differenza di quanto riscontrato nei bambini a 18 mesi, gli adulti dimostrano di possedere la capacità di apprendere e riconoscere parole nuove,

durante la fase di test, a partire da un training avvenuto in modalità unicamente visiva (col solo gesto articolatorio). Ciò significa che essi sono in grado di costruire una rappresentazione fonetica (acustica) delle parole a partire dalla conoscenza del solo gesto articolatorio ad esse associato (Havy et al., 2017).

Nel caso dei bambini di 18 mesi, parole nuove vengono apprese solo a seguito del training sonoro, che si mostra sufficiente a rendere la parola riconoscibile (e quindi appresa) sia sulla base dell'input visivo sia di quello unicamente sonoro, diversamente da quanto avviene a seguito di un training esclusivamente visivo, che, a 18 mesi, non risulta sufficiente per l'apprendimento di una parola nuova (Havy et al., 2017).

Nel caso degli adulti invece (32 studenti madrelingua inglese tra i 18 e i 22 anni) la capacità di apprendimento della parola si registra sia successivamente al training visivo sia successivamente al training esclusivamente sonoro (Havy et al., 2017).

3.3 Lettura del labiale e integrazione audiovisiva

L'input visivo influenza il processamento sonoro del linguaggio verbale già a 5 mesi (Rosenblum et al., 1997), ma, come si è visto, l'illusione chiamata effetto McGurk in cui il processo automatico di integrazione audiovisiva tra parlato e gesto articolatorio conduce a una interpretazione fallace del suono (a causa di un input articolatorio ingannevole), diventa più frequente in relazione all'aumentare dell'età degli individui (MacDonald & McGurk, 1976), (Ross et al., 2011), e

dell'esperienza linguistica (Mercure et al., 2019). In assenza di patologie, l'esperienza col linguaggio non risulta solo in una integrazione audiovisiva più marcata, ma riveste anche gli indizi articolatori di maggiore rilevanza percettiva relativamente alla dimensione fonetica, come dimostra la capacità, da adulti, di apprendere suoni linguistici nuovi a partire dal solo input articolatorio, cosa che non si verifica invece nei bambini di 18 mesi (Havy et al., 2017).

Essendo l'integrazione audiovisiva degli input del linguaggio verbale il prodotto di un processamento multisensoriale (bimodale), si ritiene utile per approfondire la natura e le implicazioni di tale processo, volgere lo sguardo sul suo funzionamento in presenza di patologie. Si è scelto di soffermarsi in prima analisi su disturbi non direttamente riconducibili a patologie del linguaggio (nella fattispecie la schizofrenia) per osservare quali ricadute possano esserci sull'integrazione audiovisiva dell'input linguistico, per poi indagare, in seconda analisi, se e come il processamento visivo del linguaggio verbale venga intaccato nel caso di disturbi connessi al linguaggio e all'apprendimento nell'età evolutiva.

A questo riguardo, si è deciso di partire dagli studi condotti su pazienti schizofrenici in cui l'esperienza linguistica è condizionata da un uso differente del linguaggio. È noto che tali individui non presentano alterazioni a livello della produzione del linguaggio (livello fonologico, semantico o sintattico). È interessante notare che dal punto di vista percettivo multimodale, le persone affette da schizofrenia sono meno soggette all'esperienza illusoria dell'effetto McGurk, seppur con variazioni individuali. Tale dato dimostra la possibilità che questo tipo di disturbo interferisca con i processi multisensoriali impliciti che servono per un'efficiente integrazione audiovisiva (White et al., 2014), ma anche

forse con i processi attentivi relativamente al processamento del volto di un parlante.

Negli individui schizofrenici l'attenzione visiva per i volti è infatti minore di quella degli individui del gruppo di controllo, ne risulta una ridotta abilità nei compiti di riconoscimento dei volti e di associazione delle emozioni corrette ai relativi volti (Loughland et al., 2002). Ai 65 individui schizofrenici, tra i 18 e i 65 anni, coinvolti nello studio erano presentati due compiti: il primo era quello di associare correttamente i volti con espressione neutra in foto nitide ai volti corrispondenti presentati in foto in cui l'immagine non era nitida; il secondo compito testava invece la capacità di associare l'emozione corretta (scegliendola tra una lista di opzioni) ai volti presentati in foto nitide (volti felici, tristi o neutri).

I risultati mostrano una minore esplorazione visiva dei volti da parte degli schizofrenici, rispetto al gruppo di controllo, e risultati peggiori nel compito di riconoscimento dell'emozione dei volti presentati, in modo particolare relativamente alle espressioni felici e neutre.

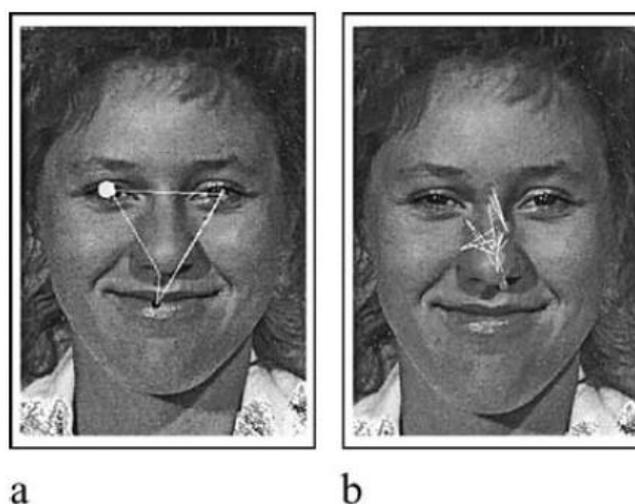


Figura 6. Risultati del tracciamento oculare di uno dei volti del test per il gruppo di individui sani (a) e il gruppo di individui schizofrenici (b) (Loughland et al., 2002).

La capacità di volgere lo sguardo alla regione del volto appropriata al momento comunicativo è in effetti correlata alla capacità di leggere efficacemente il labiale di un parlante (“lipreading” o anche “speechreading”) e alla possibilità di comprendere un numero maggiore di parole sulla sola base dell’input visivo.

Questo dato è riscontrabile già durante l’infanzia e vale sia per bambini udenti che per quelli ipoudenti (nel campione, bambini tra i 5 e gli 8 anni), i cui risultati nella comprensione delle parole presentate in modalità visiva sono superiori quando i tempi di osservazione della bocca del parlante sono maggiori (Worster et al., 2018).

Guardando ai disturbi connessi al linguaggio, si osserva che gli individui con una storia di disturbo specifico del linguaggio riportano risultati peggiori nel compito di associazione del gesto articolatorio corretto al suono di una parola sentita in modalità solo acustica rispetto al gruppo di controllo composto da individui dello stesso range di età, nello studio in oggetto, 7 – 13 anni. Si registrano inoltre potenziali evocati ridotti in relazione allo stimolo visivo del volto che articola e potenziali evocati minori relativamente alla componente N400, coinvolta nell’elaborazione semantica e registrata in risposta a casi di incongruenza semantica, alla vista di gesti articolatori incongruenti al sonoro ascoltato in precedenza (Kaganovich et al., 2016). Questo risultato suggerisce che parte dei deficit che caratterizzano il disturbo specifico del linguaggio si estende alla modalità visiva del linguaggio verbale.

Un altro studio che coinvolge 20 bambini tra i 7 e i 10 anni con disturbi del linguaggio, insieme a due gruppi, uno di adulti e uno di bambini con sviluppo tipico, riporta risultati peggiori nella lettura del labiale di parole per il gruppo con disturbi del linguaggio e performance degli adulti di gran lunga superiori rispetto a entrambi i gruppi di bambini (Heikkilä et al., 2017). Lo stesso studio rileva, sia nei bambini con disturbo specifico del linguaggio sia in quelli con sviluppo tipico, una relazione tra buone abilità fonologiche e buone capacità di lettura del labiale, una delle spiegazioni date a questo dato è che l'abilità fonologica sia probabilmente funzionale alla trasformazione dell'input visivo in informazione fonetica (Rönnerberg et al., 1998). Capacità maggiori nel compito di lettura del labiale erano inoltre correlate a migliori capacità di produzione e articolazione verbale, relativamente alla coordinazione dei movimenti articolatori, mentre i bambini che mostravano problemi nella produzione e nell'articolazione del linguaggio erano anche meno bravi nella lettura del labiale (Heikkilä et al., 2017). Dato, questo, che lascia intravedere una relazione tra capacità articolatorie e percezione visiva del linguaggio verbale.

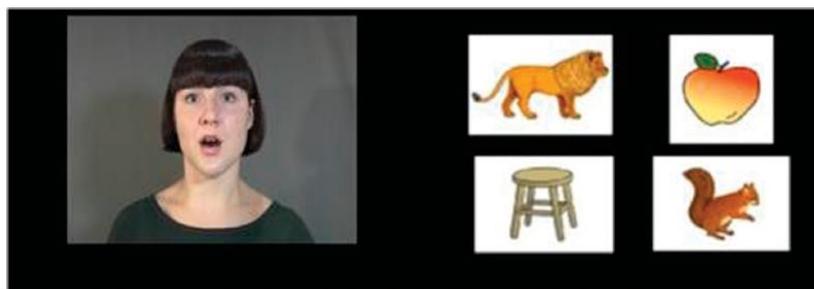


Figura 7. Fermo immagine di un video per il test di lettura del labiale. Sulla sinistra il volto che articola una parola, sulla destra la successiva presentazione delle immagini, tra le quali individuare quella della parola articolata (Heikkilä et al., 2017).

Quanto ai disturbi specifici dell'apprendimento, nello specifico la dislessia, che si manifesta con difficoltà perduranti nel tempo legate alle capacità di lettura, esistono evidenze rispetto a un processamento audiovisivo dell'input linguistico poco efficiente, che si riscontrano già durante l'infanzia e permangono in età adulta, rilevato sia a livello comportamentale (van Laarhoven et al., 2018) sia a livello neuroscientifico per gli individui che presentano questo disturbo, probabilmente legato a un deficit nel processamento dei volti e a un più generale deficit nella capacità di fare ricorso alle aree cerebrali per il processamento audiovisivo (Rüsseler et al., 2018), (Schaadt et al., 2019).

Un altro dato relativo agli individui dislessici riguarda una anomalia nella velocità di articolazione dei suoni linguistici, legata tanto al tempo effettivo di produzione quanto al tempo di pianificazione motoria del gesto linguistico durante il flusso verbale: gli individui con diagnosi di dislessia sono più lenti in entrambi i compiti rispetto al gruppo di controllo (Fawcett & Nicolson, 2002). Nello studio, veniva chiesto ai partecipanti tra i 13 e i 16 anni di pronunciare, il più velocemente e fluidamente possibile, sia ripetizioni di suoni singoli come "pu-pu-pu-pu" sia di suoni composti come "putuku". I risultati dei tempi di articolazione e di pianificazione del gesto articolatorio mostrano un deficit a entrambi i livelli: sia nella velocità di produzione, sia in quella di pianificazione motoria.

A una performance percettiva più scarsa, relativamente al processamento audiovisivo, corrisponde dunque una performance più scarsa nella produzione di linguaggio.

I dati riportati in questo paragrafo, insieme con quelli dei capitoli precedenti, permettono di spostarsi verso un livello più profondo del rapporto tra sistema articolatorio e linguaggio verbale.

Si passa dall'analisi di quella natura multimodale del linguaggio verbale che potremmo definire acustico-visiva (quindi sensoriale), resa evidente dal ruolo del gesto articolatorio quale contesto percettivo che influisce sul processamento sonoro, all'analisi di un altro aspetto della natura multimodale del linguaggio verbale: quello legato al rapporto tra componente articolatoria e produzione linguistica, ossia una multimodalità che potremmo definire acustico-articolatoria (quindi senso-motoria).

SISTEMA ARTICOLATORIO E MULTIMODALITÀ INTRINSECA

4.1 Il sistema motorio del linguaggio verbale

Fin qui si è messo in luce il ruolo del gesto articolatorio come elemento che fa del corpo del parlante un prezioso contesto percettivo. L'attenzione visiva al volto di un parlante permette infatti di arricchire l'input linguistico verbale.

Come illustrato nei paragrafi precedenti, già dai primi mesi di vita l'essere umano mostra una preferenza per i volti di parlanti in cui voce e gesto articolatorio sono sincroni e congruenti ed ha la capacità di distinguere visivamente lingue diverse, mostrando quindi l'attitudine ad attribuire caratteristiche linguistiche distintive al gesto articolatorio (cfr. paragrafi 2.1, 2.2, 2.3).

Tra i 5 e i 6 mesi l'esposizione audiovisiva all'input verbale risulta determinante al fine di distinguere fonemi acusticamente molto vicini (Teinonen et al., 2008) e a 18 mesi i bambini sono in grado di costruirsi la rappresentazione visiva del gesto articolatorio associato a una nuova parola che è stata appresa tramite il solo input sonoro (Havy et al., 2017).

Nei bilingui poi, si riscontra un incremento dell'attenzione percettiva rilevato nella capacità di distinguere visivamente lingue diverse anche a 8 mesi, cosa in

cui i monolingui della stessa età falliscono (Sebastián-Gallés et al., 2012) e, in generale, una maggiore attenzione alla regione della bocca del parlante nel primo anno di vita. I monolingui invece tendono a spostare l'attenzione alla regione della bocca quando il parlante si esprime in una lingua che non conoscono (Pons et al., 2015), fatto che denota una implicita conoscenza del valore degli indizi linguistici offerti dal gesto articolatorio.

Il ruolo del gesto articolatorio non si limita però a quello di contesto percettivo che facilita, e in certi casi, determina il corretto processamento dei suoni linguistici dati dalla voce.

La letteratura relativa alla minore efficacia nel processamento audiovisivo negli individui con storia di disturbo specifico del linguaggio e alla correlazione tra capacità di lettura del labiale e capacità di articolazione verbale negli individui con disturbo specifico del linguaggio (Kaganovich et al., 2016), (Heikkilä et al., 2017), insieme con la letteratura relativa alla maggiore lentezza articolatoria e di pianificazione motoria negli individui con diagnosi di dislessia (Fawcett & Nicolson, 2002) fornisce un primo dato significativo sul ruolo fortemente incarnato del gesto articolatorio e dunque sulla relazione tra fonazione e percezione, come illustrato nel paragrafo 3.3.

Sembra infatti che una performance percettiva più o meno efficiente si rifletta in una performance articolatoria più o meno efficiente. È dunque utile a questo punto soffermarsi sull'analisi del sistema di produzione del linguaggio verbale.

La componente biologica che consente la possibilità di articolazione linguistica nel Sapiens è il tratto vocale. La possibilità di fonazione nell'uomo è infatti data

dall'interazione tra l'apparato respiratorio (polmoni, gabbia toracica, muscoli di addome e torace), quello laringeo e quello di risonanza (vestibolo laringeo, faringe, cavità orale, cavità nasale e seni paranasali) (Brandi e Salvadori, 2004).

La meccanica della produzione verbale sfrutta gli stessi canali preposti alla respirazione e, di fatto, è nella fase di espirazione che è prodotta la voce, grazie alla trasformazione a livello laringeo dell'aria in suono, a seguito della vibrazione delle pliche (o corde) vocali.

Il suono così generato viene poi ulteriormente modulato nel passaggio attraverso il tratto vocale sopralaringeo, durante il movimento articolatorio che modifica la posizione della laringe, della lingua e delle labbra (teoria sorgente-filtro, cfr. Taylor, Reby. 2010; Falzone 2020), (Taylor & Reby, 2010), (Pennisi & Falzone, 2020).

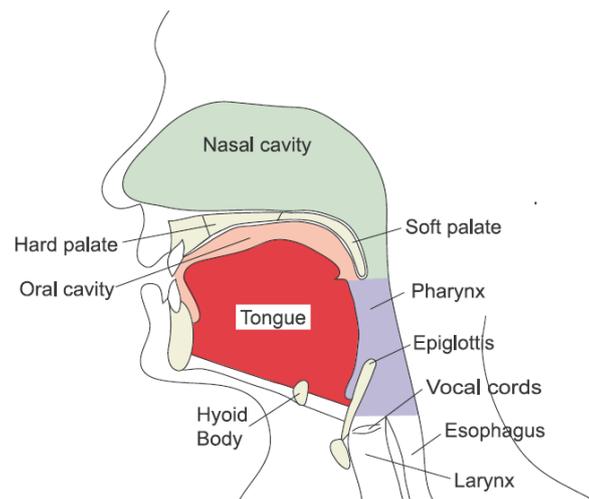


Figura 8. Tratto vocale sopralaringeo (Lieberman, P. & McCarthy, 2007).

Il tratto vocale sopralaringeo si suddivide in tratto vocale sopralaringeo verticale (SVTv) e tratto vocale sopralaringeo orizzontale (SVTh).

Il tratto vocale verticale (SVTv) è associato alla zona della gola chiamata faringe. Il tratto vocale orizzontale (SVTh) è localizzato nella cavità orale e comprende la bocca e una porzione della faringe detta orofaringe (Lieberman, P. & McCarthy, 2007).

La lunghezza delle due porzioni del tratto vocale sopralaringeo è pressoché uguale negli adulti. Tale proporzione 1:1 si raggiunge normalmente tra i 6 e gli 8 anni, permettendo fino ad allora la contemporanea ingestione di cibo e aria e preservando così dal rischio di soffocamento durante l'infanzia (Lieberman, P. & McCarthy, 2007).

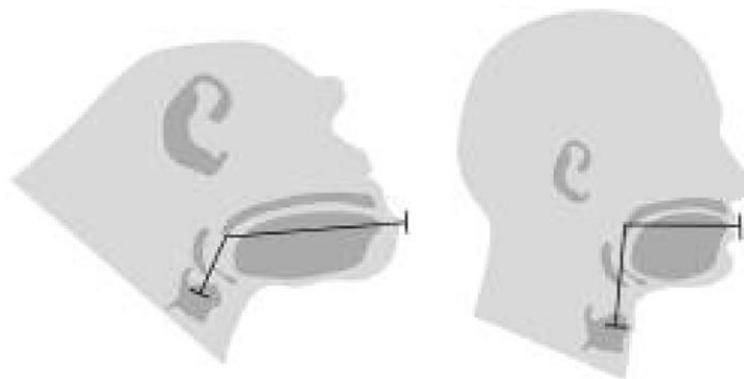


Figura 9. Esempificazione del rapporto tra SVTv e SVTh nello scimpanzé e nell'uomo (Lieberman, P. & McCarthy, 2007).

La possibilità del tratto vocale di riposizionarsi durante l'articolazione linguistica è la caratteristica che permette l'ampia gamma di suoni producibili.

I muscoli craniofacciali, laringei e linguali si caratterizzano inoltre per una composizione più idonea ai movimenti articolatori, che sono rapidissimi, e per una maggiore resistenza alla fatica se paragonati a quelli del sistema locomotore (Kent, 2004), (Ito et al., 2004). È proprio grazie a questa peculiare composizione muscolare che riusciamo a parlare lungamente senza provare un affaticamento a carico dei muscoli articolatori.

4.2 Le componenti motorie della percezione del linguaggio verbale

Il gesto articolatorio fa capo, quindi, a un sistema motorio complesso che permette l'articolazione di suoni altrettanto complessi. Il ruolo del sistema articolatorio, tuttavia, non si limita a quello di sistema atto alla produzione di linguaggio verbale.

L'attenzione visiva alla bocca del parlante è correlata a 6 mesi a una maggiore attivazione nell'area della corteccia temporale inferiore sinistra, che è coinvolta nella pianificazione dell'atto motorio articolatorio (Altvater-Mackensen & Grossmann, 2016).

Il processamento del gesto articolatorio, quando la componente sonora del linguaggio verbale è scarsa o totalmente assente, determina un'attivazione cerebrale maggiore nelle aree che rispondono agli stimoli multisensoriali nelle aree associate alla pianificazione e alla produzione linguistica. A una maggiore attivazione di queste aree corrisponde una migliore performance

nell'identificazione dell'informazione linguistica, dimostrando, per gli autori dello studio, l'importanza della simulazione interna dell'atto motorio di produzione linguistica (Callan, Jones et al., 2003).

Partendo da questi dati è più facile soffermarsi sul ruolo incarnato del gesto articolatorio, ampliandone la dimensione di sistema di produzione e di contesto percettivo finora discussa.

I neonati tra le 24 ore e i 7 giorni di vita sembrano essere già in grado di creare un'associazione motoria tra l'input verbale acustico e il movimento articolatorio corrispondente (Chen et al., 2004). Nello studio, la sperimentatrice con in braccio i neonati pronunciava i suoni /a/ e /m/.

I risultati mostrano che quando viene pronunciato il suono /a/ i neonati tendono a rispondere allo stimolo aprendo la bocca, dando una configurazione alle labbra molto simile a quella richiesta per la produzione del suono /a/. Al contrario, quando esposti allo stimolo sonoro /m/, tendono a stringere le labbra, socchiudendole, come succede per la produzione del suono /m/. Questa condizione si verifica indistintamente sia nel caso di bambini che durante l'esperimento avevano tenuto gli occhi aperti sia di quelli che avevano tenuto gli occhi chiusi, a indicare che l'associazione tra stimolo uditivo e gesto articolatorio corrispondente è in qualche misura già presente e incorporata a partire dai primi giorni di vita.



Figura 10. A sinistra immagine relativa al movimento di apertura della bocca (mouth opening). A destra immagine relativa al movimento di stretta delle labbra (mouth clutching) (Chen et al., 2004).

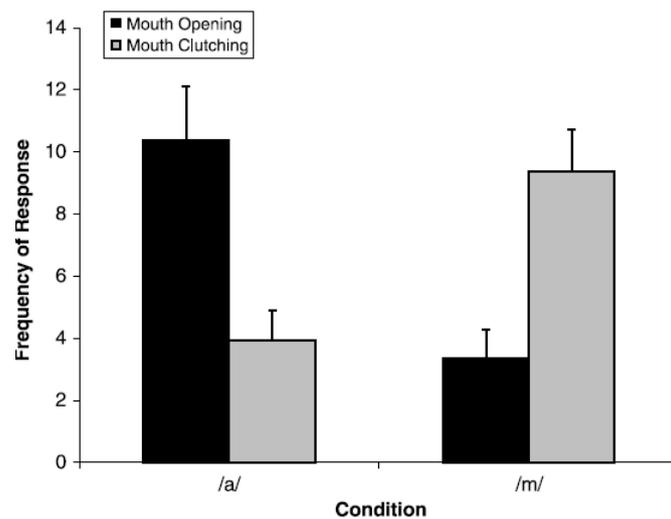


Figura 11. Frequenza media del tipo di movimento delle labbra in risposta allo stimolo sonoro ricevuto dopo 8 fasi sperimentali (Chen et al., 2004).

Un altro studio su bambini di 4 mesi e mezzo in cui i muscoli articolatori relativi alla produzione dei suoni /i/ e /u/ erano alternativamente inibiti da un giocattolo o un dito posizionati in bocca registra una forte relazione tra sistema

sensorimotorio di articolazione e percezione dei suoni linguistici (Yeung & Werker, 2013).



Figura 12. Movimento indotto nelle labbra dei bambini durante l'esperimento. In alto: il giocattolo tondo riproduce un allargamento delle labbra simile a quello necessario per la produzione vocalica /i/. In basso: il dito in bocca induce il movimento articolatorio per la produzione vocalica /u/ (Yeung & Werker, 2013).

Questo non implica però che l'esperienza visiva dell'input linguistico sia irrilevante ai fini dello sviluppo della capacità acustica e di quella articolatoria.

Al contrario, la diversa esperienza sensorimotoria del linguaggio verbale ha un'influenza sia sulla sua percezione che sulla sua produzione.

Per esempio, rispetto alle persone vedenti, i non vedenti dalla nascita, senza diagnosi di disturbi motori o del linguaggio, hanno un'acuità percettiva maggiore nel discernimento acustico di fonemi vocalici francesi presentati in un continuum

(Ménard et al., 2008); a questa aumentata acuità percettiva non corrisponde però una maggiore accuratezza articolatoria.

Relativamente alla produzione articolatoria, inoltre, il gruppo dei non vedenti, non differenzia le vocali arrotondate da quelle non arrotondate con la protrusione delle labbra in maniera tanto marcata come nel caso dei vedenti.

La differenza nel grado di protrusione delle labbra per la produzione di vocali arrotondate e vocali non arrotondate è minore nel gruppo di non vedenti, che mostrano invece, rispetto ai vedenti, riposizionamenti e cambiamenti nella forma della lingua maggiori durante la produzione di suoni vocalici arrotondati (Ménard et al., 2013) – le vocali arrotondate sono quelle la cui produzione comporta una protrusione e/o un arrotondamento delle labbra, come la “o” e la “u” per l’italiano.

Gli individui vedenti mostrano quindi nella produzione un maggiore impiego degli indizi articolatori visibili, quelli legati al movimento delle labbra. La forma e la posizione della lingua, le cui variazioni per la produzione di vocali arrotondate sono molto più ampie negli individui non vedenti, fanno sì parte del gesto articolatorio, ma sono più nascoste rispetto al movimento delle labbra.

È probabile quindi che sia la mancata esperienza visiva degli indizi articolatori visibili a creare questa differente modalità di articolazione tra vedenti e non vedenti.

A fianco di una predisposizione innata per la competenza articolatoria, sembra quindi che l’esperienza percettiva giochi a sua volta un ruolo nella costruzione e nel consolidamento di tale competenza, come dimostrerebbe la letteratura che

confronta il gesto articolatorio nella produzione linguistica di individui vedenti e non vedenti appena discussa (Ménard et al., 2008), (Ménard et al., 2013).

Allo stesso tempo, la percezione linguistica è influenzata dal sistema sensorimotorio di articolazione. Ne sono una prova gli studi che mostrano un peggioramento della performance percettiva a seguito dell'inibizione dei muscoli articolatori.

Sia durante l'ascolto del linguaggio verbale, sia durante la visione del solo movimento delle labbra relativo alla sua produzione, vengono attivate le aree cerebrali motorie coinvolte nella produzione del linguaggio (Fadiga et al., 2002), (Watkins et al., 2003), (Wilson et al. 2004), (Swaminathan et al., 2013).

In stretta relazione con questo dato, si osserva un effetto negativo nella capacità percettiva del linguaggio verbale, conseguentemente all'inibizione dei muscoli articolatori.

Uno studio condotto su bambini di 7 anni di età (età media dei partecipanti), riporta performance peggiori nel compito di lettura del labiale (senza audio) da parte dei bambini i cui muscoli articolatori erano stati inibiti chiedendo loro di stringere orizzontalmente tra le labbra un abbassalingua mentre svolgevano il compito (Turner et al., 2015).

Evidenze ancora più significative rispetto al ruolo del sistema sensorimotorio di articolazione nella percezione del linguaggio vengono dagli studi sui neonati. Nei bambini di 6 mesi, l'inibizione dei muscoli articolatori coinvolti nella produzione del suono ascoltato mediante l'uso di giocattoli per la dentizione, ha come risultato l'incapacità di distinguere tra loro due suoni consonantici non presenti

nella propria lingua madre, e quindi nuovi, ma comunque distinguibili in condizioni normali (Bruderer et al., 2015).

Lo stimolo sonoro dello studio era costituito dalla registrazione di una voce femminile che pronunciava, in maternese, due sillabe con suoni diversi della lingua hindi: il suono dentale /ḍa/ e il suono retroflesso /ɖa/. La lingua madre dei bambini dello studio era l'inglese.

A 6 mesi il restringimento percettivo di specializzazione per i suoni della propria lingua materna non è ancora avvenuto e, pur non essendo esposti a questi due suoni poiché assenti nella lingua inglese, i bambini di 6 mesi sono ancora in grado di percepirli (Werker & Tees, 2002). Ciononostante, i bambini dello studio a cui veniva dato il giocattolo piatto per la dentizione fallivano nella distinzione fonetica dei due suoni, questo perché il giocattolo di quella forma inibisce il movimento della punta della lingua, necessario per la produzione dei suoni in questione (Bruderer et al., 2015).

Uno studio successivo (Choi et al., 2019) replica i risultati ottenuti a seguito dell'inibizione dei muscoli articolatori per la produzione del contrasto fonetico /ḍa/ dentale e /ɖa/ retroflesso dell'hindi ed estende i risultati ai contrasti fonetici della lingua materna, nella fattispecie, l'inglese.

Gli input sonori dell'inglese di questi esperimenti erano il suono bilabiale /ba/ e il suono dentale alveolare /da/. I bambini ai quali veniva posizionato in bocca un giocattolo per la dentizione che impediva la chiusura delle labbra fallivano nella distinzione dei due fonemi (Choi et al., 2019).



Figura 13. (A) Immagine del giocattolo piatto, inserito nella bocca del neonato in modo da impedire il movimento della punta della lingua. (B) Immagine del giocattolo gengivale che non inibisce il movimento della lingua (Choi et al., 2019).

Questi risultati dimostrano che la capacità percettiva può essere seriamente compromessa dall'inibizione dei muscoli articolatori.

La base su cui poggia il linguaggio verbale è dunque molto più ampia di quella costituita dalla solo base uditivo-percettiva.

La configurazione articolatoria per la produzione del linguaggio verbale, come si è visto per esempio rispetto alla produzione di vocali arrotondate nei non vedenti, dipende anche dall'esperienze percettiva, ma allo stesso tempo il sistema percettivo subisce l'influenza del sistema articolatorio del linguaggio, anche in una fase della vita in cui non si è ancora fatta esperienza personale della produzione verbale e in cui le strutture articolatorie stesse non sono pienamente predisposte a consentirla.

4.3 La preferenza multimodale per il maternese

Lo sviluppo linguistico di un bambino è, come si è visto, un processo che passa attraverso una fase percettiva in cui è necessario individuare le unità strutturali della lingua a cui egli è esposto.

La dimensione comunicativa in cui normalmente sono immersi i neonati è una dimensione che sembra essere naturalmente protesa a favorire il composito processo di acquisizione del linguaggio. Questo è vero non solo perché i bambini hanno a che fare con un ambiente comunicativo che prevede quasi esclusivamente un'interazione faccia a faccia, ma anche perché chi interagisce con loro adatta il proprio stile comunicativo. È il caso del linguaggio definito infant-directed speech (IDS), a cui ci si riferisce anche con i termini motherese e baby talk, o, in italiano, maternese.

Quando interagiscono con i bambini, infatti, gli adulti tendono ad adattare automaticamente il loro registro verbale utilizzando questo registro linguistico definito maternese, che paragonato a quello normalmente usato tra adulti mostra variazioni sia nell'input sonoro sia in quello visivo del linguaggio verbale che risultano così più graditi ai neonati.

Il maternese ha delle caratteristiche prosodiche, sintattiche e lessicali che ricorrono anche tra lingue diverse, distinguendosi dal linguaggio rivolto agli adulti (adult-directed speech, da qui in poi ADS) per suoni più acuti, margini prosodici più enfaticizzati, frasi più corte, pause più lunghe, frequenti ripetizioni e un lessico più semplice (Fernald & Simon, 1984), (Grieser & Kuhl, 1988).

Queste caratteristiche sembrano rendere il maternese più attraente per le orecchie dei neonati di 4 mesi che lo preferiscono all'ascolto del registro linguistico standard adoperato nella comunicazione tra adulti (Fernald, 1985), (Cooper et al., 1997).

La preferenza dei neonati per il maternese non si esaurisce negli aspetti relativi all'input sonoro del linguaggio verbale. L'input visivo assume infatti, anche nel caso del maternese, un ruolo congiunto a quello sonoro.

Alla produzione di vocali in tono più alto rispetto all'ADS corrisponde un movimento articolatorio della bocca (la componente visibile del gesto articolatorio) più ampio e più lento durante la produzione di vocali in maternese (Kuhl et al. 1997), (Green et al., 2010) e i neonati guardano più a lungo una bocca che parla usando il maternese piuttosto che l'ADS (Lewkowicz & Hansen-Tift, 2012), (Englund & Behne, 2019).

Un altro aspetto legato al maternese riguarda l'effetto di facilitazione che esso sembra produrre nella percezione della coerenza tra i due input acustico e visivo del linguaggio verbale, legati alla produzione di frasi. La presentazione di un input linguistico sonoro, costituito da un flusso verbale continuo (non quindi ristretto a singoli suoni linguistici) e la successiva presentazione di due input visivi articolatori diversi (uno congruente all'input sonoro ascoltato, l'altro no) mostra che a 12 mesi, i bambini associano correttamente l'input sonoro ai relativi movimenti articolatori quando l'input sonoro è espresso in maternese anziché in ADS. (Kubicek, Gervain et al., 2014).

Le caratteristiche illustrate fin qui sembrano rendere il maternese particolarmente idoneo al processamento dei suoni linguistici da parte dei neonati, alle prese con il difficile compito di decodifica delle proprietà strutturali della lingua a cui sono esposti e di attribuzione delle giuste proprietà ai diversi suoni. Questo registro agevola infatti la segmentazione di parole all'interno del flusso sonoro, vale a dire la capacità di distinguere le varie parole all'interno del flusso linguistico (Thiessen et al., 2005) agendo, di fatto, come facilitatore del bootstrapping prosodico.

L'enfasi posta ai margini dei suoni semplifica la categorizzazione vocalica mentre la tendenza a produrre suoni più acuti nell'interazione adulto-bambino è probabilmente legata alla trasmissione più marcata delle emozioni che genera una maggiore attivazione dell'attenzione nel bambino (Trainor & Desjardins, 2002).

Il maternese si configura quindi come uno strumento adottato spontaneamente dagli adulti e che interviene a facilitare lo sviluppo linguistico (Kuhl et al., 1997), (Burnham et al., 2002), (Kuhl, 2004). Oltre a questo registro, è la dinamica stessa di interazione che si instaura tra caregiver e neonato a essere fortemente connotata da uno spontaneo schema di insegnamento-apprendimento.

L'interazione adulto-bambino è caratterizzata già nei primi tre mesi di vita del bambino da una risposta linguistica più frequente, da parte del caregiver, ai vocalizzi dei neonati definiti come cooing, ossia quei vocalizzi tipici dei primissimi mesi di vita, diversi dal pianto (Masataka, 2003). L'interazione è il contesto in cui si realizza la naturale didattica della comunicazione, in cui ogni

essere umano avvia il processo di acquisizione delle competenze necessarie allo sviluppo linguistico, in accordo e sotto la guida dei vincoli biologici e dei meccanismi che lo permettono.

Questo registro così peculiare viene adottato automaticamente nell'interazione con i bambini non solo dalle madri e dagli adulti quando si rivolgono ai bambini, ma persino dai bambini di 2-3 anni quando si rivolgono ai fratelli più piccoli (Dunn & Kendrick, 1982), come se la volontà di facilitare il processamento linguistico fosse una inclinazione naturale.

Esiste comunque un certo grado di flessibilità nel maternese: non si tratta infatti di un registro fisso che riflette sempre gli stessi schemi, ma al contrario, viene calibrato sulla base dell'interazione in cui si realizza.

Per esempio, sebbene esso si distingua per un maggiore utilizzo di frequenze più acute rispetto a quelle dell'ADS, si osserva la tendenza all'accrescimento dell'acutezza dei suoni nei casi in cui il bambino non reagisca allo stimolo linguistico dopo le prime due espressioni della madre (Masataka, 1992). Allo stesso tempo, il pattern sonoro delle espressioni in maternese tende a essere replicato nei vocalizzi di risposta dei neonati (Masataka, 2003).

La flessibilità e l'adattamento reciproco dell'input e della risposta che caratterizzano il maternese sembrano riflettere uno schema pedagogico che rende il maternese un registro naturalmente e intrinsecamente didattico per lo sviluppo delle capacità linguistiche e relazionali.

Il processo di acquisizione della lingua materna si realizza quindi grazie all'integrazione di vari meccanismi, specifici e generali, legati sia ai sistemi di

apprendimento sia a quelli percettivi, ed è facilitato dalla particolare, spontanea, predisposizione di un parlante più esperto a rendere più facile tale processo, come mostra l'utilizzo del maternese. Ciononostante, anche dopo l'infanzia, resta aperta la possibilità, per un essere umano, di apprendere una lingua diversa da quella materna. I paragrafi che seguono approfondiscono proprio gli aspetti legati al processo di acquisizione di una lingua diversa da quella materna, tenendo sempre come punto di partenza la natura multimodale del linguaggio verbale nella sua espressione acustico-visiva e acustico-articolatoria

CENTRALITÀ DELLA MULTIMODALITÀ DEL LINGUAGGIO NELL'APPRENDIMENTO DI UNA SECONDA LINGUA

5.1 Il processo di apprendimento di una seconda lingua

Vista la rilevanza della natura multimodale del linguaggio verbale nell'ambito dell'acquisizione e del processamento della lingua materna può essere interessante spostare l'osservazione agli effetti di tale multimodalità, data dall'integrazione audiovisiva, nell'ambito di una lingua diversa da quella materna.

La multimodalità è infatti un valore intrinseco del linguaggio verbale ed è plausibile ipotizzare che gli effetti di questa multimodalità possano essere riscontrati anche durante il processamento e l'apprendimento di una seconda lingua.

L'integrazione audiovisiva degli input verbali arricchisce il contesto percettivo, alleggerendo il compito di processamento sonoro (Van Wassenhove et al. 2005), allo stesso tempo essa può rivelarsi fondamentale nei primi mesi di vita per la possibilità di percepire contrasti fonetici molto poco marcati (Teinonen et al., 2008).

Tale effetto di facilitazione nel processamento linguistico nel contesto della lingua madre, può probabilmente mostrarsi ancora più accentuato quando si ha a che fare con suoni linguistici nuovi come quelli di una seconda lingua e quando si è maturata un'esperienza maggiore col linguaggio verbale in generale.

Così come si tende a fare maggiore ricorso agli indizi visivi del linguaggio in contesti rumorosi (Sumbly & Pollack, 1954), allo stesso modo, un input verbale nuovo può essere processato più facilmente a partire da entrambi gli input che lo caratterizzano (sonoro e visivo).

Oggi almeno metà della popolazione mondiale può essere considerata bilingue (Grosjean, 2012). Non esiste tuttavia in letteratura una definizione definitiva di "bilingue", poiché si tratta di una variabile difficilmente circoscrivibile a una serie di parametri statici utili a identificare in maniera univoca un individuo bilingue. Il bilinguismo va considerato come una variabile profondamente multidimensionale e non come una variabile esclusivamente categoriale (Luk & Bialystok, 2013).

Come sottolinea la psicologa Ellen Bialystok (2001), classificare un bilingue sulla base della sua competenza nelle lingue conosciute porta con sé controversie relative alla scelta dei parametri da utilizzare per valutare tale competenza (Bialystok, 2001).

Si può infatti stabilire un parametro che badi solo agli aspetti formali della lingua o solo a quelli comunicativi e utilizzare alternativamente un filtro che consideri come caratteristiche essenziali del linguaggio elementi differenti tra loro; in entrambi i casi è però difficile ottenere una linea di demarcazione netta al di là

della quale porre con fermezza un livello di competenza che risulti idoneo alla definizione di bilingue (nel caso in cui tale livello sia raggiunto in entrambe le lingue) e resta comunque la necessità di utilizzare l'uno o l'altro metodo, sulla base di una scelta arbitraria relativamente all'attribuzione di una maggiore attendibilità dell'uno o dell'altro.

Le dimensioni del linguaggio umano sono infatti molteplici ed è complesso stabilire all'unanimità quali siano le dimensioni fondamentali e quali invece quelle accessorie. A complicare la possibilità di una corrispondenza biunivoca tra bilinguismo e individuo bilingue ci sono ulteriori caratteristiche che rendono differenti tra loro le persone bilingue e diverso, quindi, il concetto di bilinguismo cui esse sono associate.

È molto difficile, infatti, che il bilinguismo si vada configurando in maniera identica nei diversi individui che hanno padronanza di due lingue diverse. Esso differisce non solo nel tipo di combinazione linguistica, ma anche nella modalità di appropriazione linguistica e nel tipo di esposizione alle due lingue (Romaine, 1995). Una delle distinzioni più comuni, e frequentemente oggetto di studio nelle ricerche che coinvolgono i bilingui, è quella tra bilingui precoci e bilingui tardivi. Gli stimoli ambientali occorsi nelle prime fasi dello sviluppo possono infatti incidere in maniera diversa sullo sviluppo del sistema nervoso rispetto a quelli occorsi in età adulta.

Anche all'interno di queste due sottocategorie (bilingui precoci e bilingui tardivi) vi sono comunque caratteristiche variabili a seconda dei parametri fissati. Uno dei parametri più variabili è l'età in cui ha avuto inizio l'esperienza bilingue.

Generalmente però, si considerano bilingui precoci quelli per i quali l'appropriazione linguistica è avvenuta, dalla nascita, in modo simultaneo per le due lingue e bilingui tardivi gli individui in cui l'appropriazione della seconda lingua avviene più in là nello sviluppo, tendenzialmente in età puberale o post puberale.

È proprio a partire da queste due categorie che si è fondato uno studio il cui obiettivo era quello di indagare in che modo lingue diverse venissero rappresentate nel cervello umano e di verificare le relazioni spaziali tra lingua materna e seconda lingua (Kim et al. 1997).

I due gruppi dello studio erano composti da bilingui precoci, esposti a due lingue diverse fin dall'infanzia, e bilingui tardivi, la cui età media di esposizione alla seconda lingua era poco sopra gli 11 anni. Tutti i bilingui tardivi avevano vissuto nel paese della seconda lingua e raggiunto un buon grado di padronanza linguistica.

I risultati delle osservazioni fatte con risonanza magnetica funzionale durante i compiti linguistici svolti nell'una o nell'altra lingua mostrano, nei bilingui tardivi, un'attivazione di aree distinte per la prima e la seconda lingua all'interno dell'area di Broca, l'area della corteccia frontale inferiore associata alla funzione linguistica, mentre per i bilingui precoci non si rileva una distinzione spaziale significativa, durante lo svolgimento dei compiti nelle due lingue, nelle zone di attivazione nell'area di Broca.

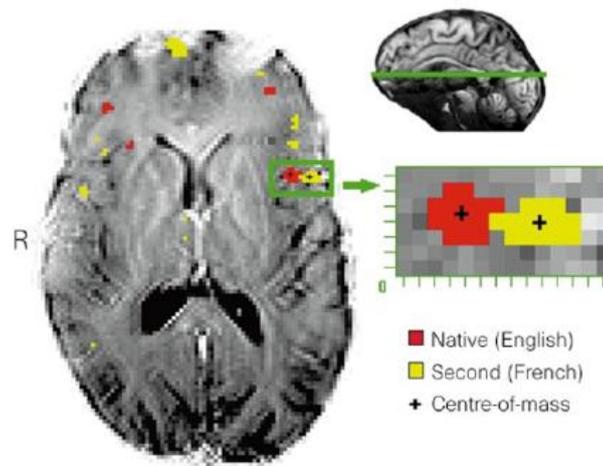


Figura 14. Sezione assiale rappresentativa dell'attivazione cerebrale nei bilingui tardivi durante il compito linguistico nelle due lingue (Kim et al., 1997).

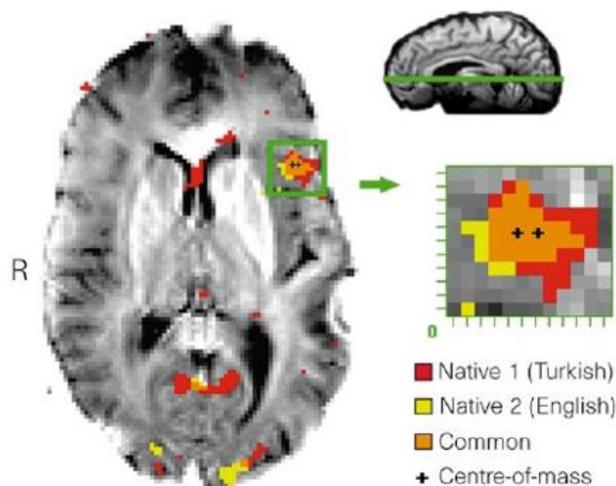


Figura 15. Sezione assiale rappresentativa dell'attivazione cerebrale nei bilingui precoci durante il compito linguistico nelle due lingue (Kim et al., 1997).

Uno studio dello stesso anno rileva una maggiore variabilità soggettiva nella rappresentazione corticale relativa al processamento della seconda lingua rispetto alla rappresentazione della prima lingua, per la quale l'attivazione

corticale risulta seguire invece un pattern pressoché fisso fra tutti gli individui (Dehaene et al., 1997).

Sebbene si possa diventare bilingui, cioè raggiungere un ottimo livello di padronanza in una lingua diversa da quella materna anche non simultaneamente all'acquisizione della prima lingua, l'esperienza di apprendimento di una seconda lingua dopo l'infanzia si configura diversamente dal processo di acquisizione della lingua madre.

Nell'ambito di questa distinzione è più corretto parlare di appropriazione linguistica, dando a questo termine la connotazione generale, riferibile indistintamente alla lingua madre o a una seconda lingua, distinguendo poi tra acquisizione nel caso della madrelingua o comunque delle lingue acquisite durante l'infanzia e apprendimento nel caso di appropriazione linguistica avvenuta in età adulta (Paradis, 2009).

Sembra quindi che l'età, o più propriamente, il periodo dello sviluppo in cui avviene l'appropriazione linguistica sia un fattore che incide in modo rilevante sui diversi meccanismi attraverso i quali avviene l'appropriazione.

Il dato della minore fluidità del processo di apprendimento linguistico se paragonato al processo di acquisizione è facilmente verificabile a livello comportamentale più o meno da chiunque abbia iniziato a studiare una seconda lingua avendo già piena padronanza della propria lingua madre, cioè approssimativamente dagli 8 anni in poi.

Ai dati che ci offre l'esperienza si affiancano diversi dati neuroscientifici. L'acquisizione linguistica sembra infatti essere più marcatamente connessa ai

meccanismi di memoria non dichiarativa, soprattutto a quelli legati alla memoria procedurale, mentre l'apprendimento linguistico (relativo quindi a una seconda lingua) è prevalentemente associato a meccanismi di memoria dichiarativa (Fabbro & Cargnelutti, 2018).

Nel caso di acquisizione di due lingue prima degli 8 anni, la rappresentazione cerebrale per gli elementi grammaticali delle due lingue è localizzata nel lobo frontale sinistro. Gli elementi semantici invece sono rappresentati nelle aree posteriori di entrambi gli emisferi (Fabbro, 2001), (Fabbro & Cargnelutti, 2018).

Nel caso di appropriazione di una seconda lingua dopo gli 8 anni gli elementi semantici delle due lingue sono cerebralmente rappresentati allo stesso modo, nelle aree posteriori dei due emisferi, mentre gli elementi grammaticali sono rappresentati nel lobo frontale sinistro per la prima lingua e nelle aree posteriori del cervello per la seconda. C'è quindi una minore rappresentazione nei sistemi di memoria procedurale per una lingua appresa dopo gli 8 anni (Fabbro, 2001), (Fabbro & Cargnelutti, 2018).

Negli individui esposti alla seconda lingua dopo i 3 anni, anche quando viene raggiunta la piena padronanza in entrambe le lingue, la rappresentazione cerebrale associata all'esecuzione di compiti grammaticali nella seconda lingua è più estesa rispetto a quella associata allo svolgimento degli stessi compiti nella prima. L'attivazione cerebrale associata allo svolgimento di compiti grammaticali nella seconda lingua cresce al crescere dell'età di prima esposizione alla seconda lingua (Wartenburger et al., 2003). Tale maggiore rappresentazione cerebrale indica un peso computazionale maggiore per il recupero delle abilità linguistiche

nella seconda lingua, nonostante l'ottimo livello di padronanza linguistica. C'è quindi un momento nello sviluppo dopo il quale le rappresentazioni di prima e seconda lingua sembrano prendere due strade diverse.

L'uomo, fin dalla nascita e indipendentemente dall'esperienza fatta durante la vita intrauterina, mostra una preferenza per i suoni linguistici (Athena Vouloumanos & Werker, 2004), (Vouloumanos & Werker, 2007a), (Vouloumanos & Werker, 2007b), ma già nei primi mesi l'esperienza linguistica occorsa in un dato momento può influenzare lo sviluppo delle capacità percettive come si è visto, per esempio, rispetto al riconoscimento visivo delle lingue da parte dei bilingui (cfr. paragrafo 2.3). I sistemi percettivi infatti attraversano dei periodi sensibili, durante i quali l'esperienza diventa determinante per il mantenersi o l'ampliarsi di alcune abilità percettive (Knudsen, 2004).

Un periodo sensibile è uno specifico periodo nello sviluppo in cui l'effetto degli input esterni è particolarmente forte. Una delle dimensioni linguistiche più legate al concetto di periodo sensibile è la dimensione fonetica, ossia la dimensione connessa ai suoni della lingua.

Nel caso della distinzione dei fonemi, il periodo sensibile è legato al concetto di restringimento percettivo.

Nei primi mesi di vita i neonati riescono a riconoscere i contrasti fonetici di tutte le lingue anche senza una precedente esposizione a quei suoni. Questa capacità percettiva va però riducendosi tra gli 8 e i 10 mesi, e la capacità di distinguere alcune differenze fonetiche viene progressivamente persa a favore di una maggiore specializzazione per i suoni della lingua cui si è esposti abitualmente

(Werker & Tees, 1984), (Polka & Werker, 1994), (Conboy & Kuhl, 2011), (Maurer & Werker, 2014).

Questo processo comporta il fatto che, man mano che si acquisisce più esperienza nella lingua a cui si è esposti, può diventare più difficile distinguere differenze fonetiche di una lingua con cui non si ha familiarità (Werker & Tees, 1983).

Di conseguenza un adulto che intende intraprendere lo studio di una seconda lingua può avere difficoltà, anche grosse, proprio nella capacità di percepire acusticamente fonemi presenti nella nuova lingua che non esistono invece nella sua lingua materna, come accade per esempio ai giapponesi con il contrasto fonetico /r/ /l/ della lingua inglese (Miyawaki et al., 1975). All'interno di questo quadro, il gesto articolatorio associato all'input sonoro può giocare un ruolo importante ai fini di una maggiore accuratezza percettiva e lo si vedrà nei capitoli successivi.

Il paragrafo che segue offre invece una panoramica relativa alle pratiche di insegnamento di una seconda lingua dalla quale emerge un progressivo e consistente impegno nel rendere la didattica delle lingue sempre più vicina alle reali dinamiche di apprendimento e alle esigenze del discente, ma non si riscontra tuttavia una particolare attenzione agli aspetti legati al carattere multisensoriale del linguaggio verbale.

5.2 La didattica della seconda lingua

Un adulto che vuole, o deve, appropriarsi di una lingua diversa da quella materna sa che dovrà fondare l'apprendimento linguistico su un processo che molto probabilmente avrà luogo in un contesto di didattica formale, che può realizzarsi nella forma di lezioni di lingua o da autodidatta, ma comunque a partire da un processo di apprendimento esplicito e consapevole.

Si chiarisce che, nell'ambito di questa trattazione, non si fa riferimento alla distinzione terminologica tra lingua seconda (da qui in avanti anche L2), intesa come lingua appresa successivamente alla lingua materna nel paese in cui si parla quella data lingua, e lingua straniera (LS), intesa come lingua appresa fuori dal paese in cui la si parla (Balboni, 1994), ma si utilizza genericamente l'espressione L2 o LS per riferirsi a una lingua appresa successivamente a quella materna (L1), quando quest'ultima è già stata consolidata.

Nel suo manuale di glottodidattica il linguista italiano Paolo Balboni (Balboni, 2012) ripercorre le tappe principali della nascita e diffusione dei diversi metodi di insegnamento delle lingue straniere, in buona parte legati ai cambiamenti ideologici di fondo o alle mutate necessità linguistiche.

Se oggi studiare lingue straniere è molto comune ed è visto sia come un'attività dilettevole di arricchimento personale sia come percorso per l'ampliamento dei propri orizzonti lavorativi e relazionali, nella Roma antica e nel corso del Medioevo e del Rinascimento lo studio di una lingua straniera era meno diffuso tra le diverse categorie sociali e avveniva prevalentemente tramite l'interazione

con un madrelingua; lo studio della grammatica era riservato alla lettura dei classici, senza alcun ricorso a libri di grammatica appositi.

È a partire dal Cinquecento e nel corso del Seicento che sorgono centri impegnati nello studio della lingua, come l'Accademia della Crusca e la Royal Society, e vedono la luce grammatiche e dizionari. Nel mondo ecclesiastico il metodo di insegnamento del latino viene esteso all'insegnamento delle lingue moderne (principalmente italiano e francese), con un approccio detto formalistico che prende forma nel Seicento ma la cui struttura, soprattutto in Italia, rimane in uso fino agli anni Sessanta.

Prendendo le mosse dalla linguistica descrittiva tradizionale, l'approccio formalistico si incentra sulla grammatica, concependo la fonologia stessa come un insieme di regole di pronuncia da imparare e mettere in pratica con esercizi di lettura. Le competenze su cui maggiormente pone l'accento l'approccio formalistico sono quelle di comunicazione scritta, con esercizi di traduzione di testi classici o di frasi create ad hoc per l'introduzione di regole ed eccezioni grammaticali. Nonostante il largo ricorso, in Italia, a un metodo formalistico di insegnamento anche negli anni Settanta-Ottanta, questo approccio negli Stati Uniti era già stato messo all'angolo nell'Ottocento, sostituito da approcci detti naturali, attenti alla dimensione psicologica e individuale del discente e a riportare al centro dell'apprendimento linguistico la dimensione orale.

L'affermarsi di una società fortemente multiculturale in America si riflette in un insegnamento linguistico che contempla, sia nelle parole del docente sia nei testi didattici proposti, non solo una trasmissione della lingua, ma anche della cultura

appartenente alla lingua, per lo sviluppo di una dimensione comunicativa più in linea con le esigenze legate a scambi commerciali e relazioni internazionali.

Il valore della dimensione orale nella didattica delle lingue viene però di nuovo svalutato a seguito degli accadimenti politici mondiali. Con lo scoppio della Prima Guerra Mondiale nel 1914 e fino alla fine del secondo conflitto mondiale il mondo è frammentato: i governi nazionalisti, in certi casi, vietano persino l'insegnamento delle lingue straniere; si viaggia meno e la lingua straniera, sul piano orale, ha meno occasioni d'uso. Diventa così centrale la lingua scritta che permette di comprendere i contenuti scientifici, letterari o informativi provenienti da altri paesi, e nasce il "Reading Method", che basa tutta la sua didattica sulla sola competenza di lettura, con testi che vengono adattati al livello del discente e gradualmente riadattati man mano che la competenza aumenta.

È però proprio durante il secondo conflitto mondiale che l'America riscopre il valore strategico delle lingue vive e a partire dalla teoria neocomportamentista dello psicologo statunitense Skinner, per il quale l'apprendimento si configura come l'esito di una serie di sequenze stimolo-risposta, e dalla teoria linguistica promossa da Bloomfield, in cui la lingua straniera è considerata a partire dalle sue unità strutturali minime, vengono messi a punto metodi didattici per le lingue che riprendono una sorta di continuità con gli approcci naturali e aggiungono l'utilizzo di nuove tecnologie, come il giradischi prima e il registratore poi, insieme talvolta a video con immagini girate nel paese della lingua oggetto di studio.

Questa nuova ottica nella didattica delle lingue straniere viene integrata dal lavoro di Robert Lado che fonda una facoltà universitaria di lingue e linguistica alla Georgetown University di Washington e promuove un approccio e una riflessione più scientificamente fondati per la didattica delle lingue. La sua linguistica contrastiva studia i rapporti e le asimmetrie tra la lingua materna e la lingua oggetto di studio, in modo da anticipare i nodi problematici e mettere in atto strategie di insegnamento più adeguate.

L'interazione tra la teoria di Skinner, la linguistica di Bloomfield e l'analisi contrastiva di Lado dà vita ai cosiddetti metodi strutturalistici, in cui assumono un ruolo centrale gli esercizi di ripetizione con feedback, finalizzati alla memorizzazione di elementi linguistici e all'automatizzazione di alcuni processi, con un utilizzo sempre più ampio di tecnologie come i registratori – che all'input linguistico del docente madrelingua fanno seguire la possibilità di registrare la propria produzione linguistica, per poi verificarne la correttezza riascoltando l'interpretazione del madrelingua.

La fine degli anni Cinquanta vede il diffondersi della teoria di Chomsky (1959), che mette in discussione il modello comportamentista di apprendimento; allo stesso tempo il pensiero di Lado va evolvendosi nella direzione della sociolinguistica per la quale l'approccio strutturalistico perde di valore poiché nessuna delle microstrutture linguistiche oggetto di studio ha senso fuori dal contesto sociale che sta alla base della comunicazione. Negli anni Sessanta pubblicazioni come *How to Do Things with Words* (Austin, 1962), del filosofo e linguista John Austin e *Speech acts* (Searle, 1969) di John Searle portano al centro

il ruolo pragmatico della lingua. Dire qualcosa non è una pura descrizione del fare ma consiste invece proprio nel fare quella cosa.

A partire da queste idee prende forma l'approccio comunicativo all'insegnamento delle lingue straniere. La lingua è vista come un mezzo per l'azione sociale e si fa inoltre strada la nozione di competenza comunicativa che è il vero obiettivo dell'appropriazione linguistica e ingloba al suo interno la competenza linguistica, le competenze extralinguistiche e le competenze contestuali. Sulla scia di questo approccio prendono forma vari metodi didattici come il metodo nozionale-funzionale, che si radica fortemente nella pragmalinguistica e nella sociolinguistica, e la glottodidattica umanista, in cui le dimensioni psicologica e neuropsicologica diventano sempre più centrali e si tengono in considerazione anche gli aspetti emozionali del discente.

Sotto l'ombrello dell'approccio comunicativo si colloca anche il metodo naturale del linguista californiano Stephen Krashen. Nell'illustrare la distinzione tra acquisizione di una lingua come processo inconscio e apprendimento di una lingua come processo consapevole, Krashen spiega quali siano i percorsi da seguire perché il processo di appropriazione di una seconda lingua negli adulti proceda per acquisizione e non per apprendimento, facendo ricorso allo stesso Language Acquisition Device di impronta chomskyana cui hanno accesso i bambini per l'acquisizione della lingua materna (Krashen, 1982).

L'acquisizione, mediata dall'insegnante, deve avvenire partendo da un input comprensibile all'interno di un contesto di interazione significativa. Tale input deve collocarsi al gradino subito successivo rispetto alle conoscenze già

possedute (k+1), in un ordine naturale in cui ogni input è successivo al precedente, con un'attenzione al filtro affettivo che può inibire l'acquisizione in quanto meccanismo di autodifesa legato a sensazioni di ansia rispetto alle proprie capacità o alla capacità di risolvere compiti non in linea con l'ordine naturale di acquisizione. Allo stesso modo, la correzione dell'errore risulta superflua perché poco funzionale all'acquisizione e più legata a un processo di apprendimento esplicito. L'errore è visto come un passaggio necessario e indicativo dell'acquisizione linguistica, così come lo è per i bambini alle prese con la lingua materna, e non come un errore nello svolgimento di un compito.

Il concetto di ordine naturale di acquisizione è oggetto di studio della linguistica acquisizionale che si concentra sulla nozione di interlingua (Selinker, 1972), ossia la lingua che usa chi sta imparando una nuova lingua, e dalla quale è possibile riconoscere gli stadi di acquisizione, anche a partire dagli errori commessi. L'applicazione di una regola grammaticale, anche nei casi che fanno eccezione a quella data regola (per esempio la formazione del passato dei verbi irregolari fatta applicando la regola generale), non indica una scarsa abilità linguistica bensì l'acquisizione di una regola che viene, di fatto, generalizzata anche ai casi che fanno eccezione. Selinker (1972) osserva che il corpus dinamico dell'interlingua ha delle caratteristiche generali comuni a tutte le interlingue come, per esempio, l'omissione dei morfemi grammaticali, quelle desinenze delle parole che danno di esse un'informazione grammaticale, come la "i" del plurale in italiano.

Balboni (2012) osserva che l'ipotesi acquisizionale relativamente all'appropriazione di una lingua una volta che si possiede già padronanza della lingua materna non trova grandi riscontri in ambito didattico poiché il discente

che ha già piena conoscenza di una lingua parte da questa per costruire un processo consapevole di apprendimento e non di acquisizione.

La tendenza globale della glottodidattica contemporanea verte verso un approccio comunicativo, ma tende a conservare anche aspetti metodologici più vecchi in una mescolanza di metodi e teorie di fondo mirata a rendere quanto più efficace l'apprendimento linguistico.

Si riscontra tuttavia, nell'ambito della didattica delle lingue, una carenza di considerazione della matrice multisensoriale del linguaggio verbale. La competenza nell'espressione orale costituisce uno degli aspetti più difficili per il discente e soffermarsi sugli aspetti intrinseci del linguaggio verbale può invece essere utile per semplificare il processo di acquisizione delle competenze orali in una seconda lingua.

5.3 Gesto articolatorio e seconda lingua

Alla luce dei dati raccolti e tenendo al centro la multimodalità del linguaggio verbale, è possibile orientare l'osservazione verso il ruolo del gesto articolatorio nel processo di apprendimento di una seconda lingua.

Il fatto che i bambini nei primi mesi di vita facciano affidamento anche sugli indizi visivi offerti dalla bocca del parlante e il fatto che anche in età adulta tali elementi possano influenzare o facilitare la percezione dei suoni (MacDonald & McGurk, 1976), (Sumbly & Pollack, 1954), insieme con il fatto che i bilingui simultanei si focalizzano maggiormente sugli indizi offerti dal gesto articolatorio di un parlante

(Pons et al., 2015), crea le fondamenta su cui poggiare l'ipotesi che il gesto articolatorio possa configurarsi non solo come un prezioso contesto percettivo per la lingua materna, ma anche come elemento facilitatore per l'apprendimento della seconda lingua.

Come per il maternese (Kuhl et al., 1997), (Burnham et al., 2002), (Kuhl, 2004) anche nel linguaggio usato per rivolgersi a una persona di cui si riconosce la scarsa competenza nella lingua di interazione si è riscontrata una tendenza all'iperarticolazione vocalica (Uther et al., 2007).

L'iperarticolazione delle vocali nel maternese è vista sotto la luce di un suo possibile ruolo didattico nella facilitazione del riconoscimento dei suoni linguistici (Kuhl et al., 1997), (Kuhl, 2004).

Il dato relativo all'aggiustamento del registro linguistico da parte del madrelingua nell'interazione con persone non madrelingua, potrebbe quindi essere considerato a sua volta come una sorta di adattamento naturale che riflette una conoscenza implicita rispetto al valore della componente articolatoria nel processamento linguistico (Uther et al., 2007).

Navarra e Soto-Faraco (2007) hanno condotto uno studio per verificare se gli indizi articolatori contribuiscano al miglioramento delle abilità di riconoscimento degli elementi fonetici in una L2 (Navarra & Soto-Faraco, 2007).

I monolingui spagnoli hanno difficoltà nel percepire la differenza sonora dei fonemi /ε/ e /e/ tipici della lingua catalana, categorizzandoli nell'unico fonema /e/ della lingua spagnola (Pallier et al., 1997), (Navarra et al., 2005).

I due autori hanno quindi indagato le differenze nelle capacità di riconoscimento dei due fonemi di individui con spagnolo lingua dominante e con catalano lingua dominante, nella condizione unicamente sonora, unicamente visiva e nella condizione audiovisiva.

I partecipanti allo studio erano tutti bilingui, 103 studenti universitari, la cui coppia linguistica era spagnolo-catalano. 50 di loro erano nati in famiglie monolingui catalano (catalano lingua dominante) e 53 in famiglie monolingui spagnolo (spagnolo lingua dominante). Tutti i partecipanti erano stati esposti alla loro seconda lingua (quella non dominante della coppia linguistica) entro i 3 anni, generalmente al nido o alla materna. Il campione venne diviso in due gruppi: uno la cui lingua dominante era il catalano e il secondo la cui lingua dominante era lo spagnolo.

I risultati mostrano che nella condizione solo sonora gli individui del gruppo spagnolo lingua dominante non sono in grado di riconoscere la differenza fonetica, mentre i bilingui con catalano lingua dominante, nella medesima condizione sonora, distinguono i due fonemi /ɛ/ e /e/. Nella condizione di presentazione unicamente visiva invece, nessuno dei due gruppi distingue i due fonemi.

Nella condizione audiovisiva, nella quale l'input sonoro è integrato con il correlato visivo del gesto articolatorio, entrambi i gruppi distinguono i due fonemi. E questo nonostante il fatto che per la presentazione audiovisiva non venisse in alcun modo esplicitata l'istruzione di prestare attenzione alla componente articolatoria del suono (Navarra & Soto-Faraco, 2007).

Il gesto articolatorio si configura in questo caso come elemento determinante per la percezione del contrasto fonetico per il gruppo con lo spagnolo come lingua dominante.



Figura 16. Istantanea dei video presentati nello studio del momento di massima apertura della bocca durante l'articolazione dei fonemi catalani /ε/ (a sinistra) e /e/ (destra) (Navarra & Soto-Faraco, 2007).

Un altro studio (Hirata & Kelly, 2010), che si muove nella stessa direzione, si concentra sulle differenze quantitative di durata del fonema, anziché qualitative, come nel caso dello studio precedente. L'obiettivo è di valutare l'impatto di un training audiovisivo paragonando i risultati ottenuti a quelli ricavati da un training unicamente sonoro.

Entrambi i training erano mirati a migliorare la capacità di percepire la differenza tra vocali lunghe e vocali corte in giapponese.

Sebbene entrambi i gruppi testati riportarono un miglioramento dalla fase di pre test a quella post test, il training audiovisivo, rispetto al training unicamente sonoro, risultò avere un impatto significativamente maggiore sull'acquisizione dell'abilità di distinguere vocali lunghe da vocali corte (Hirata & Kelly, 2010).

Il restringimento percettivo e la progressiva specializzazione per i suoni della lingua a cui si è esposti nel corso dello sviluppo hanno un effetto sull'abilità dei giapponesi di percepire la differenza tra i suoni /r/ e /l/ dell'inglese (Miyawaki et al., 1975). Questa differenziazione fonemica non esiste in lingua giapponese e risulta infatti particolarmente difficile da percepire e da produrre per i giapponesi che studiano inglese.

A seguito di un training della durata di 3-4 settimane, mirato all'apprendimento dell'identificazione percettiva dei fonemi /r/ e /l/, le performance di riconoscimento del contrasto in questione migliorano sia per le parole presentate durante la fase di training, sia per parole nuove, mostrando un effetto di generalizzazione del miglioramento percettivo come conseguenza del training (Bradlow et al., 1997).

L'effetto del training sembra essere preservato nel tempo (dopo 3 mesi e dopo 6 mesi), anche in assenza di training di mantenimento (Lively et al., 1994); lo stesso training ha effetti positivi sull'abilità di produzione di parole contenenti i suoni /r/ e /l/.

Nonostante non fosse prevista nella fase di training alcuna parte dedicata a questo aspetto, a seguito del training si registra comunque anche un miglioramento delle performance, valutate da madrelingua inglesi, per la produzione delle parole con i due fonemi (Bradlow et al., 1997).

I dati riportati nello studio sono tutti estratti da analisi comportamentali e non dicono molto su quali possano essere gli effetti a livello cerebrale di un training

mirato al miglioramento delle performance per il riconoscimento di contrasti fonetici difficili di una seconda lingua.

È proprio con l'obiettivo di individuare le aree cerebrali coinvolte nell'apprendimento percettivo di difficili contrasti fonetici di una seconda lingua che si struttura lo studio condotto da Daniel Callan e i suoi colleghi (Callan, Tajima, et al., 2003).

Il contrasto fonetico preso in esame è ancora quello della lingua inglese /r/ - /l/, come si è visto estremamente difficile da percepire per i parlanti di lingua giapponese.

Le persone coinvolte nello studio erano tutte di madrelingua giapponese, tra i 21 e i 30 anni, con all'attivo 6 anni di classi di lingua inglese in ambito scolastico, focalizzate sulle competenze di lettura e scrittura e meno sulla capacità di espressione orale. L'insegnamento dell'inglese era comunque avvenuto in Giappone e nessuno dei partecipanti aveva mai vissuto in un paese anglofono.

L'osservazione dell'attività cerebrale nello studio è condotta con risonanza magnetica funzionale sia nella fase precedente al training percettivo, svolto nell'arco di un mese, sia nella fase successiva al suo svolgimento.

Nella fase precedente al training si registra una maggiore attività cerebrale nelle aree connesse al sistema motorio in risposta al processamento del contrasto fonetico inglese /r/ - /l/ rispetto a quella per il contrasto fonetico /b/ - /g/, che costituisce una coppia fonemica della lingua inglese facile da distinguere anche per i giapponesi.

I risultati comportamentali, dopo il training per l'identificazione percettiva dei fonemi /r/ e /l/, mostrano un miglioramento delle prestazioni nel compito di riconoscimento dei due fonemi.

I risultati di neuroimaging mostrano un incremento dell'attività cerebrale connessa al processamento delle parole contenenti i due fonemi nella fase successiva al training, quando le capacità percettive risultano affinate, in linea con gli studi che rilevano una maggiore estensione dell'attività cerebrale per la seconda lingua (Wartenburger et al., 2003).

L'incremento e la riorganizzazione dell'attività cerebrale riguarda sia le aree già attive durante il processamento sonoro antecedente al training sia l'attivazione di aree cerebrali nuove. A riorganizzarsi sono le aree temporali superiori e mediali connesse con il processamento acustico-fonetico.

Le nuove aree di attivazione comprendono invece aree corticali e sottocorticali legate alla mappatura uditivo-articolatoria collegata al processamento e all'apprendimento del linguaggio verbale.

Nello specifico si osserva un'attivazione bilaterale nella circonvoluzione sopramarginale, nel planum temporale, nell'area di Broca, nella corteccia motoria supplementare e nella corteccia premotoria, e un'attivazione in aree sottocorticali quali il cervelletto, i gangli della base e la sostanza nera.

Gli autori concludono che la maggiore capacità percettiva, a seguito del training, sia il risultato dell'acquisizione di mappe percettivo-motorie che rendono la percezione più facile, perché compiuta a partire da azioni articolatorie potenziali (Callan, Tajima, et al., 2003).

Anche durante il solo ascolto del linguaggio si osserva un'attivazione bilaterale di una porzione superiore della corteccia premotoria ventrale, che coincide in buona parte con alcune delle aree coinvolte nella produzione di linguaggio (Wilson et al. 2004). Tuttavia, la rappresentazione cerebrale motoria sembra essere maggiore quando si ascoltano contrasti fonetici di una lingua che non è la propria lingua materna e quando si ascoltano contrasti fonetici della propria lingua madre prodotti da parlanti di un'altra lingua e, quindi, pronunciati con un accento con cui si ha poca familiarità (Callan et al., 2014).

Il fatto che il training audiovisivo negli studi riportati (Navarra & Soto-Faraco, 2007), (Hirata & Kelly, 2010) dia risultati percettivi migliori rispetto al training unicamente sonoro si pone a mio avviso in linea con l'interpretazione riportata negli studi di Callan e colleghi (Callan, Jones et al., 2003), (Callan, Tajima, et al., 2003).

È possibile, infatti, che il training audiovisivo faciliti la capacità percettiva perché l'integrazione audiovisiva di gesto articolatorio e input sonoro costituisce una fonte esplicita, seppur non completa, di quella mappa articolatorio-uditiva del suono linguistico che, se acquisita, migliora le capacità percettive e di produzione di contrasti fonetici difficili.

CONCLUSIONI

6.1 La doppia natura multimodale del linguaggio verbale

L'ipotesi che il processamento multisensoriale (audiovisivo) del linguaggio verbale possa facilitare non solo l'acquisizione della prima lingua, ma anche l'apprendimento di una seconda lingua in un'età diversa da quella di acquisizione della facoltà di linguaggio, sembra trovare un riscontro nei dati che riportano migliori capacità percettive per i suoni di una L2 a seguito di un training audiovisivo, in cui lo stimolo sonoro è presentato in associazione al gesto articolatorio correlato (Hirata & Kelly, 2010), (Navarra & Soto-Faraco, 2007).

Questo effetto di facilitazione per la L2, dato dall'integrazione audiovisiva dei due input del linguaggio verbale, ha però una natura più profonda, che va oltre il semplice avvalersi di uno stimolo multisensoriale per facilitare il processamento dell'input e che può essere rintracciata a partire dalle basi motorie del linguaggio.

Il processamento di suoni linguistici attiva aree cerebrali preposte al processamento uditivo, ma comporta al contempo l'attivazione di alcune delle

aree coinvolte nella produzione di linguaggio, come la parte superiore della corteccia premotoria ventrale (Wilson et al. 2004).

Il dato relativo alla più estesa attivazione delle aree motorie del linguaggio durante l'ascolto di suoni di una lingua straniera (Wilson & Iacoboni, 2006) e l'incremento di questo tipo di attivazione per l'ascolto dei fonemi più difficili di una lingua diversa da quella materna (Callan, Tajima et al., 2003), (Callan et al., 2004), (Callan et al., 2014) segnalano un maggior coinvolgimento del sistema motorio, a livello cerebrale, durante il processamento di una lingua straniera.

Sembra quindi che il maggiore ricorso alle aree motorie del linguaggio sia associato a stimoli sonori con cui si ha minore dimestichezza.

È, di nuovo, rispetto alle aree motorie del linguaggio che si registra un'attivazione maggiore per il processamento di fonemi di una lingua straniera per i quali non si ha un corrispettivo articolatorio e sonoro nella propria lingua madre (Schmitz et al., 2019). È il caso delle vocali tedesche /ö/ e /y/ per parlanti di lingua italiana, per le quali l'attivazione motoria è maggiore se paragonata all'ascolto di vocali tedesche con un corrispettivo articolatorio e acustico in lingua italiana; segno che il sistema motorio, in questi casi, interviene in qualche misura a supplire alla mancanza di esperienza sensorimotoria (uditiva e articolatoria) rispetto a un dato fonema (Schmitz et al., 2019).

A fronte di questi dati, è interessante osservare che a seguito di un training che migliora le capacità percettive (e di produzione) di un contrasto fonetico difficile di una seconda lingua, come quello costituito dai fonemi /r/ e /l/ per i giapponesi alle prese con l'inglese, si registrano, in risposta al processamento dei suoni in

questione, un incremento e una riorganizzazione dell'attività in alcune aree cerebrali.

Le nuove aree di attivazione si rintracciano in aree corticali e sottocorticali associate alla mappatura uditivo-articolatoria dei suoni linguistici (Callan, Tajima et al., 2003): ciò significa che il sistema motorio entra in gioco in modo cruciale (ma non indipendente), per facilitare il compito di riconoscimento di suoni linguistici poco familiari.

La maggiore attivazione motoria e il suo crescere al crescere del livello di difficoltà di distinzione acustica e al grado di esperienza sensorimotoria che si ha di determinati fonemi pongono la componente motoria del linguaggio in una posizione di centralità quando si ha a che fare con una seconda lingua.

La necessità di stabilire nuove mappe uditivo-articolatorie per i suoni che appartengono a un'altra lingua spiega probabilmente questo maggiore ricorso alla componente motoria del linguaggio, chiamato in causa proprio quando la componente sonora, da sola, non è sufficiente all'identificazione dei fonemi.

Lì dove l'onda sonora da sola non basta, è il sistema motorio a intervenire per riportare l'esperienza acustica a una dimensione uditivo-articolatoria che permetta di processare e identificare con più accuratezza il suono linguistico.

Il linguaggio verbale assume dunque una dimensione multimodale costituita da due livelli: il primo legato alla sua natura multisensoriale corporea, che può, nel caso della comunicazione vis-à-vis, contribuire ad arricchire il contesto percettivo in cui collocare il suono linguistico; il secondo, legato al rapporto che il suono linguistico, quando lo si processa, instaura con il gesto articolatorio

potenziale, individuabile nell'attivazione delle aree cerebrali motorie legate alla produzione linguistica.

La maggiore attivazione delle aree motorie per l'identificazione di contrasti fonetici difficili di una lingua straniera e la riorganizzazione, conseguente a un training che migliora le capacità percettive e produttive di quei suoni, in aree cerebrali connesse alla creazione di mappe uditivo-articolatorie relative al suono (Callan, Tajima et al., 2003) riflettono proprio questo rapporto.

6.2 La multimodalità del linguaggio verbale e l'apprendimento di una seconda lingua

I due livelli di multimodalità del linguaggio verbale (acustico-visiva e acustico-articolatoria) enucleati nel paragrafo precedente sono distinti: il maggiore ricorso alle aree cerebrali motorie del linguaggio durante il processamento di una lingua straniera o di contrasti fonetici più impegnativi è infatti indipendente dalla visione del gesto articolatorio.

Esiste però un punto, in superficie, in cui i due livelli convergono e a partire dal quale può esserci un effetto di facilitazione nel processo di appropriazione di una seconda lingua. Questo punto di congiunzione risiede nella componente articolatoria visibile, vale a dire il movimento articolatorio della bocca, quella parte del gesto articolatorio che costituisce l'input visivo a partire dal quale il linguaggio verbale può essere processato in modo audiovisivo, integrando gesto articolatorio e input sonoro.

Questa integrazione multisensoriale automatica, relativamente al linguaggio verbale, è evidente già nei primi mesi di vita (Dodd, 1979), (Kuhl & Meltzoff, 1982), (Rosenblum et al., 1997), (Patterson & Werker, 2003), (Denis Burnham & Dodd, 2004). Il legame dell'integrazione audiovisiva per il linguaggio verbale diventa più forte nel corso dello sviluppo (Ross et al., 2011), (Mercure et al., 2019), comportando, negli adulti, una maggiore incidenza dell'effetto di percezione illusoria di un suono a causa di un indizio articolatorio fuorviante (MacDonald & McGurk, 1976).

Al di là di casi sperimentali ingannevoli, l'integrazione audiovisiva degli input verbali risulta vantaggiosa, perché arricchisce il contesto percettivo in cui processare l'input sonoro, alleggerendo così il processo computazionale (Brancazio & Miller, 2005), (Van Wassenhove et al. 2005). Si mostra inoltre estremamente preziosa quando si interagisce in contesti in cui il rumore di fondo interferisce con la percezione sonora (Sumby & Pollack, 1954), situazione in cui si tende automaticamente a portare l'attenzione visiva verso la bocca del parlante, proprio perché essa fornisce indizi utili alla ricostruzione sonora.

La maggiore esperienza col linguaggio si riflette anche nella capacità di crearsi una rappresentazione fonetica di una parola nuova a partire dal solo gesto articolatorio. Gli adulti sono infatti in grado di riconoscere acusticamente parole nuove a cui sono stati esposti solo in modalità visiva, cosa di cui i bambini di 18 mesi non sono capaci (Havy et al., 2017).

Pure qualora si ammetta che l'integrazione audiovisiva del linguaggio verbale non si fonda sul sistema motorio, come mostrano alcuni studi sperimentali in cui

l'inibizione articolatoria non comporta una riduzione dell'effetto McGurk, come ci si potrebbe aspettare se le basi del processo di integrazione audiovisiva fossero motorie (Matchin et al. 2014), non si può negare che l'esperienza audiovisiva abbia un impatto sulla componente articolatoria e che essa incida sulla costruzione di mappe possibili di relazione uditivo-articolatoria e che, quindi, essa possa agire come elemento di facilitazione per l'acquisizione di nuove mappe uditivo-articolatorie di una seconda lingua.

È chiaro che il processamento audiovisivo dell'input linguistico non è una condizione indispensabile per lo sviluppo della capacità articolatoria: non si spiegherebbe altrimenti la piena capacità articolatoria acquisita dai non vedenti dalla nascita. Però, la minore accuratezza articolatoria (Ménard et al., 2008) e la differente conformazione del gesto articolatorio per la produzione, per esempio, di vocali arrotondate (come la "o" e la "u" per l'italiano) da parte dei non vedenti, (Ménard et al., 2013) sembrano essere una spia del fatto che il gesto articolatorio visibile possa contenere delle informazioni che vengono assimilate e integrate nella propria esperienza di produzione linguistica. La protrusione delle labbra tipica della produzione di vocali arrotondate, per esempio, è minore negli individui non vedenti dalla nascita, che per la produzione di questi suoni si avvalgono invece di maggiori riposizionamenti e variazioni nella forma della lingua (le parti nascoste del gesto articolatorio) rispetto ai vedenti.

Sebbene il gesto articolatorio visibile non mostri tutti gli indizi articolatori, è plausibile supporre, alla luce dei dati raccolti, che esso informi e modelli, in parte, la produzione linguistica di chi ha accesso a questo input.

In quest'ottica, la possibilità di partire da un input audiovisivo per i contrasti fonetici più impegnativi di una lingua straniera può rappresentare una scorciatoia per l'accesso alla costituzione di quelle mappe uditivo-articolatorie che migliorano la percezione e la produzione di quei suoni linguistici e che dipendono dall'attivazione consistente delle aree motorie del linguaggio.

Si spiegherebbero quindi nel rapporto che intercorre tra i due livelli di multimodalità del linguaggio verbale (acustico-visiva e acustico-articolatoria) i migliori risultati percettivi a seguito dell'esposizione audiovisiva a un input linguistico impegnativo di una seconda lingua, rispetto ai risultati ottenuti con l'esposizione unicamente sonora, poiché il gesto articolatorio fa da ponte tra i due livelli di multimodalità, mettendoli in relazione l'uno con l'altro.

L'input visivo del gesto articolatorio, infatti, può veicolare informazioni rispetto alla combinazione articolatorio-uditiva dei suoni ascoltati, facilitando così il compito di creazione o di ricorso alle mappe uditivo-articolatorie che permettono una più proficua percezione sonora e una migliore produzione linguistica.

Il crescere dell'età alla quale si inizia a studiare una lingua diversa da quella materna sembra avere un impatto positivo sui compiti linguistici che richiedono abilità metalinguistiche e analitiche, ma costituisce invece uno svantaggio per l'acquisizione delle competenze di comprensione orale (Muñoz, 2006), tuttavia, il punto di contatto dei due livelli di multimodalità del linguaggio verbale (acustico-visiva e acustico-articolatoria) offerto dal gesto articolatorio visibile, potrebbe, in qualche misura, capovolgere il paradigma per il quale apprendere una lingua, soprattutto relativamente ai suoi aspetti fonologici, è inevitabilmente

più difficile in relazione all'avanzare dell'età in cui ci si appresta a farlo, o perlomeno può porre l'accento su un elemento a cui l'adulto può fare ricorso con una certa garanzia di profitto e che fa capo a una dimensione propria anche dell'acquisizione della lingua materna in cui il gesto articolatorio facilita la percezione di contrasti fonetici acusticamente molto vicini (Teinonen et al., 2008).

Negli adulti, infatti, il processo di integrazione audiovisiva è più marcato (MacDonald & McGurk, 1976), (Brancazio & Miller, 2005), (Van Wassenhove et al. 2005) e la capacità di inferire informazioni fonetiche dal gesto articolatorio è maggiore rispetto a quella dei bambini (Baart et al. 2014), (Havy et al., 2017). Quindi, forse, proprio gli adulti possono trarre maggiore vantaggio dall'esposizione audiovisiva ai suoni di una seconda lingua.

Sarebbero utili, in relazione a questa affermazione, studi sperimentali che paragonino i risultati di adulti e bambini nel compito di identificazione fonetica per i suoni di una L2 successivamente a un training audiovisivo, ma è plausibile che l'esperienza linguistica (e quindi l'età) renda più proficua l'estrazione di informazioni fonetiche dal gesto articolatorio, come già dimostrato da studi condotti nell'ambito della prima lingua (Baart et al. 2014).

Le migliori capacità di percezione delle differenze fonetiche di suoni non presenti nella propria lingua materna, riscontrate durante l'esposizione audiovisiva all'input, anziché solo sonora (Navarra & Soto-Faraco, 2007), e l'incremento della capacità di percezione nella differenza di durata dei dittonghi di una lingua straniera a seguito di un training audiovisivo (Hirata & Kelly, 2010) anziché

unicamente sonoro, creano dunque un ponte tra le riflessioni fatte sulla doppia natura multimodale del linguaggio verbale (acustico-visiva e acustico-articolatoria) e la didattica delle lingue.

In primo luogo, rispetto al ruolo circoscritto che può assumere un'esercitazione specifica condotta a partire dall'input audiovisivo per un contrasto fonetico difficile o per l'identificazione delle differenze nella durata dei dittonghi.

In secondo luogo, in una visione più ampia, rispetto all'importanza che il contesto naturale di esposizione linguistica multimodale assume in quest'ottica nell'apprendimento di una seconda lingua e all'effetto di facilitazione che produce la visione del gesto articolatorio degli input verbali.

L'attenzione alla regione della bocca di un parlante nei primi mesi di vita è correlata al successivo sviluppo linguistico (Young et al., 2009), (Tenenbaum et al., 2015), allo stesso tempo tale attenzione è a sua volta modulata, già nel primo anno di vita, dal tipo di esperienza linguistica che si fa (Pons et al., 2015). I bilingui, bambini esposti fin dalla nascita a lingue diverse, mostrano nel primo anno di vita una maggiore attenzione per la regione della bocca di chi parla (Pons et al., 2015), differenziandosi in questo dai monolingui che mostrano invece una distribuzione più bilanciata tra attenzione selettiva agli occhi e alla bocca del parlante (Lewkowicz & Hansen-Tift, 2012).

Contestualmente, nei bilingui si registra un incremento dell'attenzione percettiva rispetto alla capacità di distinguere lingue diverse a partire dalla sola esposizione al gesto articolatorio (senza la voce) del parlante (Sebastián-Gallés et al., 2012).

Se, intorno agli 8 mesi, i monolingui non riescono più a distinguere visivamente lingue diverse, i bilingui riescono invece a farlo anche relativamente a lingue che non appartengono alla loro coppia linguistica e di cui, quindi, non hanno fatto esperienza percettiva (Sebastián-Gallés et al., 2012).

Questi elementi supportano l'ipotesi che il gesto articolatorio visibile rappresenti un fulcro importante per il processamento dell'input verbale e che esso acquisisca automaticamente maggiore salienza percettiva quando si ha a che fare con lingue meno familiari, tanto che, nonostante i monolingui a 12 mesi guardino in egual misura gli occhi e la bocca del parlante, essi portano l'attenzione in misura maggiore verso la regione della bocca quando il parlante utilizza una lingua che non conoscono (Pons et al., 2015).

Allo stesso tempo, però, il restringimento percettivo che si rileva intorno agli 8 mesi rispetto all'incapacità dei monolingui di distinguere visivamente la propria lingua materna da un'altra (Sebastián-Gallés et al., 2012) sembra essere recuperato in età adulta, quando i monolingui, seppur con un livello di precisione minore rispetto ai bilingui, sono in grado di distinguere visivamente la propria lingua da un'altra (Soto-Faraco et al., 2007). Dato, questo, che avvalorava l'idea che al crescere dell'esperienza, cresca anche la capacità di inferire informazioni linguistiche dal gesto articolatorio.

Questi dati insieme mostrano che il ricorso, nella computazione cerebrale, all'elemento articolatorio del linguaggio verbale varia al variare della richiesta del compito linguistico e si affina con l'aumentare dell'esperienza che si ha del linguaggio.

Il gesto articolatorio visibile permea il linguaggio verbale nel ruolo di anello di congiunzione tra la sua dimensione acustico-visiva e quella acustico-articolatoria. Esso, quando disponibile, viene infatti processato automaticamente, insieme all'input sonoro.

Tuttavia, la sua rilevanza percettiva non è costante, ma si adatta sulla base del peso della computazione sonora. Ne è un esempio il fatto che l'identificazione del contrasto fonetico /ɛ/ ed /e/ del catalano sia migliore durante l'esposizione audiovisiva all'input, invece che solo sonora, seppur non vi sia alcuna istruzione esplicita rispetto all'attenzione da porre sul gesto articolatorio (Navarra & Soto-Faraco, 2007).

Il fatto, poi, che a una migliore capacità di identificazione fonetica corrisponda una migliorata capacità produttiva (Bradlow et al., 1997) risulta particolarmente interessante per il processo di apprendimento di una L2.

La multimodalità, nei suoi due livelli qui enucleati, rappresenta quindi una risorsa intrinseca del linguaggio verbale e non una risorsa ristretta al contesto della lingua materna.

Lo sviluppo della capacità articolatoria, ove non vi siano patologie specifiche, è garantita dal nostro sistema articolatorio anche in assenza dell'input visivo del verbale, ma al tempo stesso l'esperienza visiva del gesto articolatorio può influenzare la conformazione articolatoria dei suoni linguistici (Ménard et al., 2008), (Ménard et al., 2013) e facilitarne la comprensione sonora (Navarra & Soto-Faraco, 2007), (Teinonen et al., 2008), (Hirata & Kelly, 2010).

Sebbene l'integrazione audiovisiva sia un processo automatico e riscontrabile già nei neonati (Dodd, 1979), (Kuhl & Meltzoff, 1982), (Rosenblum et al., 1997), (Patterson & Werker, 2003), (Burnham & Dodd, 2004), la possibilità di accedere all'input audiovisivo del linguaggio verbale diventa particolarmente preziosa da adulti, alle prese con una lingua poco familiare.

Se si pensa alla didattica delle lingue è importante porre al centro l'importanza di una risorsa così biologicamente connaturata come la multimodalità del linguaggio verbale e non trattare l'esperienza di apprendimento come il passaggio di un corpus linguistico che può essere astratto dal corpo che lo produce. Questo non significa necessariamente contemplare unicamente una didattica dal vivo, ma avvalersi, sempre, della chiave multimodale che permea l'espressione orale, anche quando l'apprendimento è affidato a supporti tecnologici.

Nell'ambito degli studi sulla L2, si parte generalmente dall'assunto che il suo apprendimento si fondi su percorsi diversi da quelli utilizzati per l'acquisizione della lingua materna, e in effetti più studi confermano che l'età in cui ha luogo l'appropriazione linguistica influenza la sua rappresentazione a livello cerebrale (Kim et al. 1997), (Dehaene et al., 1997), (Fabbro 2001), (Wartenburger et al., 2003), (Fabbro & Cargnelutti, 2018).

La ricostruzione qui proposta, tuttavia, fa emergere una caratteristica automatica di facilitazione per l'appropriazione linguistica, rilevabile già a 5 mesi (Teinonen et al., 2008). Questa caratteristica non viene persa nel corso dello sviluppo, ma, al contrario, vede aumentare la capacità di renderla funzionale all'apprendimento

con l'aumentare dell'esperienza linguistica. L'integrazione audiovisiva dell'input verbale è un processo automatico, che non dipende dalla lingua che si processa o dalla volontà, o meno, di avviarlo. È invece una componente intrinseca del processamento del linguaggio verbale vis-à-vis che facilita la possibilità di apprendere dall'input, così come succede di norma per gli stimoli multisensoriali (Hatwell, 2003), (Bremner et al. 2012).

L'importanza della dimensione multimodale del linguaggio verbale nello sviluppo del linguaggio e nell'apprendimento di una seconda lingua, e il ruolo chiave del gesto articolatorio visibile nel creare un punto di contatto tra il livello acustico-visivo e quello acustico-articolatorio di questa multimodalità, attribuiscono una connotazione corporea all'onda sonora prodotta col linguaggio, che diventa in quest'ottica molto più di una comune alterazione nella densità delle molecole dell'aria. Essa porta al suo interno un'informazione corporea che è funzionale tanto alla sua percezione quanto alla sua produzione.

6.3 Applicazioni pratiche

Come si è visto nel corso di questa trattazione, la bocca di un parlante e il gesto articolatorio visibile che essa produce durante l'atto linguistico (verbale), rappresenta un elemento importante sia per lo sviluppo della lingua materna, sia per il processamento e la produzione di una seconda lingua.

Ricordiamo brevemente, per esempio, che a tempi di osservazione maggiori della regione della bocca del parlante nei bambini di 6 mesi, corrisponde uno sviluppo

maggiore del linguaggio espressivo a 18 mesi (Young et al., 2009) e che, a 6 mesi, maggiore è il tempo in cui i bambini guardano la bocca che articola i suoni linguistici, maggiore è la riproduzione dei suoni (Imafuku et al., 2019). Si è riscontrato inoltre che i bambini a 4 mesi e mezzo riproducono più a lungo i suoni vocalici emessi dal volto di un parlante in cui gesto articolatorio e input sonoro sono concordanti (Patterson & Werker, 1999) e che la vista del gesto articolatorio può diventare determinante per la capacità, tra i 5 e i 6 mesi, di distinguere fonemi acusticamente molto vicini (Teinonen et al., 2008).

L'integrazione audiovisiva dell'input verbale, anziché il suo processamento unicamente sonoro, si mostra a sua volta estremamente utile ai fini della percezione, in età adulta, di fonemi non presenti nella propria lingua materna (Navarra & Soto-Faraco, 2007), (Hirata & Kelly, 2010). Ciononostante, la rilevanza del gesto articolatorio ai fini di un più accurato processamento linguistico non sembra essere un dato preso in grande considerazione nell'ambito della didattica di una seconda lingua. Si è visto infatti che proprio nel contesto del processamento dei suoni di una seconda lingua, a una migliore capacità di percezione dei suoni linguistici, corrisponde un miglioramento della capacità di riproduzione di quei suoni (Bradlow et al., 1997), (Callan, Tajima et al., 2003) ed è in questo contesto, quindi, che avvalersi del processamento audiovisivo dell'input verbale diventa particolarmente importante.

Da un'analisi di tre delle applicazioni più utilizzate per l'apprendimento delle lingue straniere disponibili per Android, "Duolingo" (più di 100 milioni di download), "Babbel" (più di 10 milioni di download) e "Busuu" (più di 10 milioni di download), emerge una totale assenza di considerazione della dimensione

multimodale (acustico-visiva e acustico-articolatoria) del linguaggio verbale: gli esercizi di ascolto prevedono la presentazione dell'input sonoro e l'associazione di tale input al corrispettivo scritto o la possibilità di registrare una propria versione delle parole ascoltate, a partire dal solo input sonoro, ricevendo poi un feedback grafico sulla qualità della propria pronuncia.

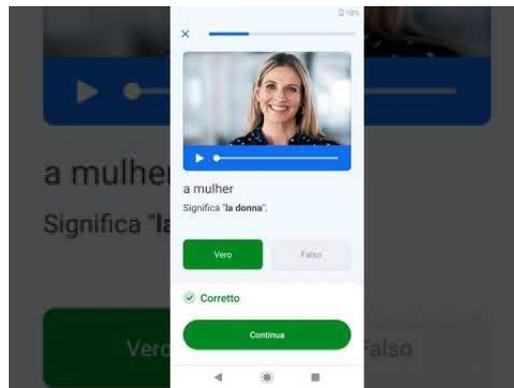
I suoni linguistici sono dunque presentati a partire dal solo input sonoro, scorporati da tutto e associati visivamente, al massimo, ai caratteri scritti. È vero che le applicazioni per i cellulari, pur promettendo buoni risultati con una pratica quotidiana ristretta anche a 5 soli minuti al giorno, non hanno probabilmente la pretesa di porsi come corsi esemplari di didattica delle lingue, tuttavia la loro diffusione tende ad alimentare la considerazione del linguaggio verbale come prodotto unisensoriale, con il rafforzamento del suo carattere uditivo e il risultato di un totale annullamento della sua natura intrinsecamente multimodale.

Si riportano di seguito tre estratti di esercizi per l'apprendimento del lessico tratti rispettivamente dall'App Busuu per la lingua portoghese brasiliano (video 2), dall'App Babbel per la lingua danese (video 3) e dall'App Duolingo per la lingua russa (video 4).

La scelta delle tre lingue, portoghese, danese e russo è stata fatta per offrire uno scenario più ampio dell'effetto della presentazione di un input linguistico nella sola modalità sonora: per il lettore italiano, generalmente, il portoghese ha dei suoni più comprensibili ed è maggiore la possibilità che ci sia stato qualche contatto preliminare con la lingua (per esempio con canzoni brasiliane celebri anche in Italia). Nel caso del danese, per un italiano, la conoscenza preliminare

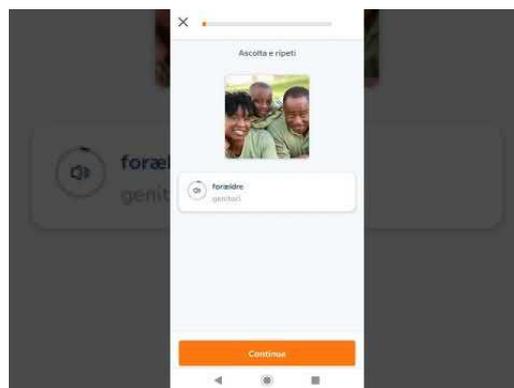
degli aspetti fonetici della lingua, è molto probabilmente, più scarsa e c'è inoltre una minore vicinanza con i suoni linguistici dell'italiano. Infine, la lingua russa, di cui sia i suoni linguistici che i caratteri dell'alfabeto sono distanti da quelli della lingua italiana.

Fare click sull'immagine per avviare il video



Video 2. App Busuu, esercizio di lessico per la lingua portoghese.

Fare click sull'immagine per avviare il video



Video 3. App Babbel, esercizio di lessico per la lingua danese.

Fare click sull'immagine per avviare il video



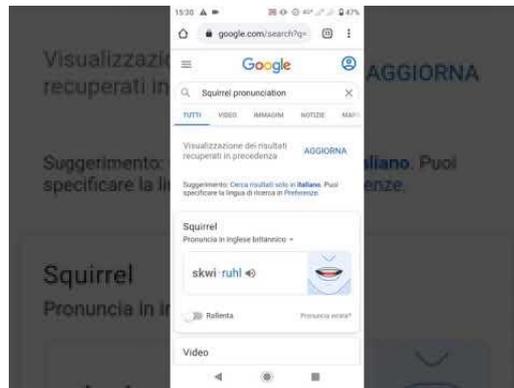
Video 4. App Duolingo, esercizio di lessico per la lingua russa.

La comprensione orale e, ancor di più, la produzione orale in una lingua straniera sono sempre indicate dagli studenti come la parte più difficile del processo di apprendimento. Spesso, a buone capacità di comprensione dei testi scritti si associano capacità piuttosto scarse di comprensione e produzione orale. Per questo motivo è fondamentale dedicare grande attenzione alla creazione dei contenuti che nascono per sviluppare tali competenze.

Lo strumento gratuito di Google per l'ascolto della pronuncia di parole in lingua straniera mostra una, seppur basilare, considerazione per la natura multimodale e multisensoriale dell'input verbale, affiancando al suono riprodotto una rappresentazione grafica schematizzata del relativo gesto articolatorio, con la possibilità ulteriore di riprodurre suono e gesto articolatorio a una velocità ridotta.

Di seguito viene riportato un esempio di questa rappresentazione acustico-articolatoria creata e messa a disposizione gratuitamente da Google. La parola pronunciata è “squirrel”, ossia scoiattolo, in lingua inglese (video 5).

Fare click sull’immagine per avviare il video



Video 5. Tool di pronuncia di Google.

Il fatto di lavorare su interfacce digitali non rappresenta un limite alla possibilità di presentare l’input sonoro in associazione al gesto articolatorio che lo produce con una rappresentazione più vicina a quella naturale (e non solo abbozzata come quella vista nello strumento di Google).

Al contrario, i progressi tecnologici e grafici permettono di ricreare in maniera estremamente accurata il gesto articolatorio associato alla produzione verbale insieme con il relativo input sonoro: il tutor animato Baldi ne è un esempio (Massaro & Palmer Jr, 1998). Si tratta di un’interfaccia grafica per computer costituita da una testa animata che produce linguaggio e offre la possibilità di mostrare non solo la vista frontale del volto che pronuncia una parola con tanto di gesto articolatorio visibile, ma anche rappresentazioni laterali e interne che illustrano i dettagli del gesto articolatorio interno, quello normalmente non visibile, su cui si fonda la produzione linguistica.

L'efficacia del training audiovisivo per mezzo dell'ausilio di tale interfaccia è stata testata relativamente all'identificazione fonetica di parole inglesi contenenti i fonemi /r/ e /l/, da parte di individui giapponesi. Oltre al compito di identificazione fonetica si è poi valutata la capacità di produzione linguistica successiva al training (Massaro & Light, 2003).

La fase di training prevedeva due modalità di presentazione dell'input sonoro: una associata alla sola presentazione standard del volto frontale che articola la parola congruentemente alla riproduzione del suono ad essa relativo (da qui in avanti condizione A); l'altra con la presentazione di 4 diverse rappresentazioni del gesto articolatorio, comprensive della parte normalmente invisibile relativa all'articolazione interna oltre che di quella visibile (da qui in avanti condizione B). Nello specifico, le presentazioni visive della condizione B includono una rappresentazione frontale con la pelle del volto resa trasparente, una posteriore con vista dell'interno, una riproduzione laterale della bocca (in modalità statica e dinamica) e una riproduzione integrale del volto visto lateralmente con la pelle trasparente.

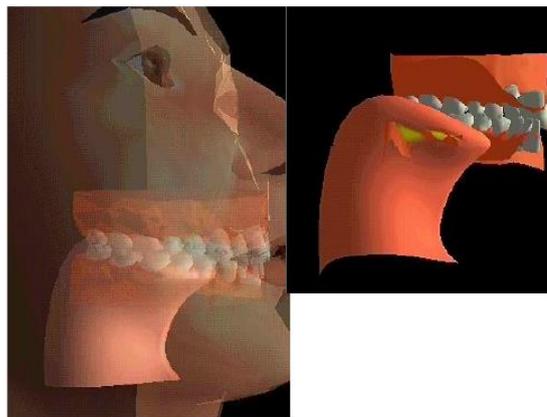


Figura 17. Esempi di vista laterale degli articolatori interni nella condizione B (Massaro & Light, 2003).

In entrambe le modalità di training, Baldi forniva istruzioni dettagliate su come posizionare la lingua per riprodurre il suono ascoltato e la velocità di produzione linguistica era ridotta del 63% rispetto alla velocità normale.

L'analisi dei risultati rivela un miglioramento nella performance di identificazione fonetica e di produzione linguistica conseguente a entrambe le condizioni di training, ma non riporta, relativamente al training della condizione B, un vantaggio significativo in nessuno dei due compiti (identificazione fonetica e produzione), nonostante la più completa presentazione del gesto articolatorio (comprensiva della visione interna).

Al termine dello studio i partecipanti riportavano un feedback positivo rispetto alla possibilità offerta nella condizione B di osservare gli indizi articolatori interni, ciononostante pare proprio che l'aggiunta della vista di tali elementi non costituisca un elemento più funzionale ai fini dell'apprendimento linguistico.

Questo dato si pone in linea con la proposta che il gesto articolatorio visibile, pur non mostrando la configurazione completa dell'atto articolatorio, rappresenti il punto in cui convergono i due livelli di multimodalità del linguaggio verbale, quella acustico-visiva e quella acustico-articolatoria, e a partire dal quale si ha un effetto di facilitazione nella percezione fonetica che si riflette poi in un miglioramento della capacità di produzione proprio per effetto del ricorso alla multimodalità doppia del linguaggio verbale (cfr. paragrafi 6.1 e 6.2). Il gesto articolatorio visibile era infatti presentato in entrambe le condizioni di training con l'interfaccia Baldi.

Dunque, lì dove la rappresentazione e l'esperienza fonologica di un suono sono carenti, come nel caso dei suoni di una lingua straniera, l'attivazione motoria diventa più consistente ed è in quest'ottica che il volto di un parlante esperto nella produzione di quei suoni può diventare cruciale nell'offrire un elemento che faccia da ponte tra i due livelli di multimodalità del linguaggio verbale, facilitando il recupero o la creazione di nuove mappe uditivo-articolatorie che permettano di processare più accuratamente un suono linguistico migliorando così la capacità di produzione dello stesso.

Esiste quindi una risorsa, biologicamente determinata che non perde, ma, al contrario, vede crescere il suo potere nel passaggio dalla prima alla seconda lingua, e alla quale si può attingere per facilitare il processo di appropriazione linguistica.

Questa risorsa biologica così preziosa, costituita dal gesto articolatorio visibile, non ha necessariamente bisogno di eleganti e complesse riproduzioni grafiche (che sono comunque utili e funzionali) ma è reperibile molto facilmente semplicemente guardando la persona che parla, che diventa in quest'ottica essa stessa una risorsa.

Come afferma più poeticamente Mariangela Gualtieri (2010), "la vita ha bisogno di un corpo per essere e tu sii dolce con ogni corpo [...]. Sia placido questo nostro esserci, questo essere corpi scelti per l'incastro dei compagni d'amore", dove per compagni d'amore si intendono qui tutti i conspecifici il cui valore, relativamente alla relazione tra individuo esperto e individuo inesperto, attribuisce valore a noi stessi (che alterniamo costantemente il nostro esistere tra la condizione di

esperti e quella di inesperti) e quindi, auspicabilmente, maggiore senso di cura per la nostra specie e per l'ecosistema in cui essa si colloca.

Fare click sull'immagine per avviare il video



Video 6. Mariangela Gualtieri recita la sua poesia "mio vero" da cui è preso l'estratto proposto nella parte conclusiva della trattazione (Gualtieri, 2010).

Riferimenti bibliografici

- Acredolo, L. P., Goodwyn, S. W. (1985). Symbolic gesturing in language development. *Human Development, 28*(1), 40–49.
- Altvater-Mackensen, N., & Grossmann, T. (2016). The role of left inferior frontal cortex during audiovisual speech perception in infants. *NeuroImage, 133*, 14–20.
<https://doi.org/10.1016/j.neuroimage.2016.02.061>
- Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press.
- Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition, 130*(1), 31-43.
- Balboni, P. E. (1994). *Didattica dell'italiano a stranieri*. Bonacci.
- Balboni, P. E. (2012). *Le Sfide di Babele. Insegnare le lingue nelle società complesse*. Torino: UTET Università.
- Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology, 7*(2), 113–127.
- Bates, E. (1979). *The emergence of symbols*. Academic Press.
- Bialystok, E. (2001). *Bilingualism in Development: Language, literacy, and cognition*. Cambridge University Press.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310.
- Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception and Psychophysics, 67*(5), 759–769. <https://doi.org/10.3758/BF03193531>
- Brandi, Luciana e Salvadori, B. (2004). *Dal suono alla parola: Percezione e produzione del linguaggio tra neurolinguistica e psicolinguistica* (F. U. Press, ed.).

- Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). Multisensory Development. In *Multisensory Development*.
<https://doi.org/10.1093/acprof:oso/9780199586059.001.0001>
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*. <https://doi.org/10.1162/jocn.2009.21076>
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536. <https://doi.org/10.1073/pnas.1508631112>
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435-1435.
- Burnham, Denis, & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*. <https://doi.org/10.1002/dev.20032>
- Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience*, *8*, 275.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. *NeuroImage*, *22*(3), 1182-1194.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, *19*(1), 113–124.
- Callan, Daniel E., Jones, J. A., Munhall, K., Callan, A. M., Callan, A. M., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*, *14*(17), 746–748. <https://doi.org/10.1097/00001756-200312020-00016>

- Campisi, E. (2018). *Che cos'è la gestualità*. Carocci. Carrocci editore.
- Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, 7(1), 42–47. <https://doi.org/10.1111/j.1467-7687.2004.00321.x>
- Choi, D., Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al.(2015). *Psychonomic Bulletin & Review*, 26(4), 1388–1399.
- Choi, Dawoon, Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). *Psychonomic Bulletin and Review*, 26(4), 1388–1399. <https://doi.org/10.3758/s13423-019-01601-0>
- Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, 35(1), 26-58.
- Church, R. B., & Goldin-Meadow, S. (2017). So how does gesture function in speaking, communication, and thinking. In *Why Gesture* (pp. 387–412).
- Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first-and second-language speech perception. *Developmental Science*, 14(2), 242-248.
- Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*.
[https://doi.org/10.1016/S0163-6383\(97\)90037-0](https://doi.org/10.1016/S0163-6383(97)90037-0)
- Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton University Press.
- de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16(1), 1–17.
- de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11(1), 1-15.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176.

- Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, 370(6487), 292–295.
<https://doi.org/10.1038/370292a0>
- Dehaene-lambertz, G., Dehaene, S., & Hertz-pannier, L. (2013). *Functional Neuroimaging of Speech Perception in Infants*. 298(December 2002), 2013–2016.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *European Review*, 16(4), 399–411. <https://doi.org/10.1017/S1062798708000513>
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., ... Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*. <https://doi.org/10.1097/00001756-199712010-00030>
- Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*. [https://doi.org/10.1016/0010-0285\(79\)90021-5](https://doi.org/10.1016/0010-0285(79)90021-5)
- Dodd, B. E., & Campbell, R. E. (1987). *Hearing by eye: The psychology of lip-reading*. Lawrence Erlbaum Associates, Inc.
- Dunn, J., & Kendrick, C. (1982). The speech of two- and three-year-olds to infant siblings: 'Baby talk' and the context of communication. *Journal of Child Language*, 9(3), 579–595.
<https://doi.org/10.1017/S030500090000492X>
- Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4(1), 40–48. <https://doi.org/10.1111/1467-7687.00147>
- Endress, A. D., Nespors, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, 13(8), 348–353.
<https://doi.org/10.1016/j.tics.2009.05.005>
- Englund, N., & Behne, D. M. (2019). Perception of audiovisual infant directed speech. *Scandinavian Journal of Psychology*, 61(2), 218–226. <https://doi.org/10.1111/sjop.12599>
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. <https://doi.org/10.1038/415429a>

- Fabbro, F. (2001). The Bilingual Brain: Cerebral Representation of Languages. *Brain and Language*, 79(2), 211–222. <https://doi.org/10.1006/brln.2001.2481>
- Fabbro, F., & Cargnelutti, E. (2018). *Neuroscienze del Bilinguismo. Il farsi e il disfarsi delle lingue*. Roma: Astrolabio.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399–402. <https://doi.org/10.1046/j.0953-816x.2001.01874.x>
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.152159999>
- Fawcett, A. J., & Nicolson, R. I. (2002). Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia*. <https://doi.org/10.1002/dys.222>
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*. [https://doi.org/10.1016/S0163-6383\(85\)80005-9](https://doi.org/10.1016/S0163-6383(85)80005-9)
- Fernald, A., & Simon, T. (1984). Expanded Intonation Contours in Mothers' Speech to Newborns. *Developmental Psychology*, 20(1), 104–113.
- Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespors, M., & Mehler, J. (2016). On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain. *Developmental Science*, 19(3), 488–503. <https://doi.org/10.1111/desc.12323>
- Feyereisen, P. (1987). Gestures and Speech, Interactions and Separations: A Reply to McNeill (1985). *Psychological Review*. <https://doi.org/10.1037/0033-295X.94.4.493>
- Fishbane, P. M., Gasiorowicz, S. G., Thornton, S. T. S. (1993). *Physics for Scientists and Engineers: Extended Version*. Englewood Cliffs, New Jersey: 1377 S. Prentice Hall.
- Fontana, S. (2008). Mouth actions as gesture in sign language. *Gesture*, 8(1).
- Fontana, S. (2009). *Linguaggio e multimodalità: gestualità e oralità nelle lingue vocali e nelle lingue dei segni*. ETS.

- Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: Humanlike pattern of Wernicke's brain language area homolog. *Science*, 279(5348), 220–222. <https://doi.org/10.1126/science.279.5348.220>
- Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. In *SSRN*. <https://doi.org/10.1146/annurev.psych.093008.100408>
- Gibson, J. J., & Pick, A. D. (1963). Perception of another person's looking behavior. *The American Journal of Psychology*. <https://doi.org/10.2307/1419779>
- Gleitman, L. R., & Landau, B. (1994). *The acquisition of the lexicon*. MIT Press.
- Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, 1((1)), 3–55.
- Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development*. <https://doi.org/10.1111/1467-8624.00138>
- Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, 10(6), 778–785. <https://doi.org/10.1111/j.1467-7687.2007.00636.x>
- Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*.
- Grieser, D. L., & Kuhl, P. K. (1988). *Maternal Speech to Infants in a Tonal Language : Support for Universal Prosodic Features in Motherese*. 24(1), 14–20.
- Grosjean, F. (2012). *Bilingual. Life and reality*. Cambridge, MA: Harvard.
- Gualtieri, M. (2010). *Bestia di gioia*. Einaudi.
- Guasti, M. T. (2007). *L'acquisizione del linguaggio. Un'introduzione*.
- Hains, S. M. J., & Muir, D. W. (1996). Infant Sensitivity to Adult Eye Direction. *Child Development*, 67(5), 1940–1951. <https://doi.org/10.1111/j.1467-8624.1996.tb01836.x>
- Hatwell, Y. (2003). Intermodal coordinations in children. Touching for knowing. *Intermodal Coordinations in Children. Touching for Knowing: Cognitive Psychology of Haptic Manual*

Perception, 53, 207.

- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*.
[https://doi.org/10.1016/S0010-0277\(00\)00132-3](https://doi.org/10.1016/S0010-0277(00)00132-3)
- Havy, M., Foroud, A., Fais, L., & Werker, J. F. (2017). The Role of Auditory and Visual Speech in Word Learning at 18 Months and in Adulthood. *Child Development*, 88(6), 2043–2059.
<https://doi.org/10.1111/cdev.12715>
- Hayes, B. (1995). Metrical Stress Theory: Principles and Case Studies. In *Chicago University of Chicago Press*.
- Heikkilä, J., Lonka, E., Ahola, S., Meronen, A., & Tiippana, K. (2017). Lipreading ability and its cognitive correlates in typically developing children and children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 60(3), 485–493.
https://doi.org/10.1044/2016_JSLHR-S-15-0071
- Hirata, Y., & Kelly, S. D. (2010). Effects of Lips and Hands on Auditory Learning of Second-Language Speech Sounds. *Journal of Speech, Language, and Hearing Research*.
[https://doi.org/10.1044/1092-4388\(2009/08-0243\)](https://doi.org/10.1044/1092-4388(2009/08-0243))
- Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science*, 22(6), 1–12.
<https://doi.org/10.1111/desc.12825>
- Ito, T., Murano, E. Z., & Gomi, H. (2004). Fast force-generation dynamics of human articulatory muscles. *Journal of Applied Physiology*, 96(6), 2318–2324.
<https://doi.org/10.1152/jappphysiol.01048.2003>
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–371.
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, 37(3), 967–979.
- Kaganovich, N., Schumaker, J., & Rowland, C. (2016). Atypical audiovisual word processing in school-age children with a history of specific language impairment: An event-related

potential study. *Journal of Neurodevelopmental Disorders*, 8(1), 1–22.

<https://doi.org/10.1186/s11689-016-9168-3>

Kent, R. D. (2004). The uniqueness of speech among motor systems. *Clinical Linguistics and Phonetics*, 18(6–8), 495–505. <https://doi.org/10.1080/02699200410001703600>

Kim, K. H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*. <https://doi.org/10.1038/40623>

Kimura, D. (1999). *Sex and cognition*. MIT press. MIT press.

Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412–1425. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.1663&rep=rep1&type=pdf>
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.467.1663>

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, 40(5), 419–435. <https://doi.org/10.1006/jhev.2001.0468>

Krashen, S. (1982). *Principles and practice in second language acquisition*.

Kubicek, C., Gervain, J., De Boisferon, A. H., Pascalis, O., Lœvenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior and Development*, 37(4), 644–651.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684–686.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*. <https://doi.org/10.1098/rstb.2013.0302>

- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1114783109>
- Lieberman, P., & McCarthy, R. (2007). Tracking the evolution of language and speech : comparing vocal tracts to identify speech capabilities. *Expedition*, 49(2), 15–20.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Long-term Retention of New Phonetic Categories. The Journal of the Acoustical Society of America*, 96(4), 2076–2087.
- Loughland, C. M., Williams, L. M., & Gordon, E. (2002). Schizophrenia and affective disorder show different visual scanning behavior for faces: A trait versus state-based distinction? *Biological Psychiatry*. [https://doi.org/10.1016/S0006-3223\(02\)01356-2](https://doi.org/10.1016/S0006-3223(02)01356-2)
- Low, E. B., & McCarthy, J. M. (1983). A Theory of Internal Reduplication. *Linguistic Review*. <https://doi.org/10.1515/tlir.1983.3.1.25>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*. <https://doi.org/10.1080/20445911.2013.795574>
- Maccoby, E. E. (1966). *The development of sex differences*. Stanford University Press.
- MacDonald, H., & McGurk, J. (1976). Hearing lips and seeing voices. *Nature*, 764. <https://doi.org/10.1038/260170a0>
- Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, 53(2), 155–180. [https://doi.org/10.1016/0010-0277\(94\)90069-8](https://doi.org/10.1016/0010-0277(94)90069-8)
- Masataka, N. (1992). Pitch Characteristics Of Japanese Maternal Speech To Infants. *Journal of Child Language*, 19(2), 213–223. <https://doi.org/10.1017/S0305000900011399>
- Masataka, N. (2003). The onset of language. In *The Onset of Language*. <https://doi.org/10.1017/CBO9780511489754>

- Massaro, D. W., & Light, J. (2003). Read my tongue movements: bimodal learning to perceive and produce non-native speech/r/and/l. *Eighth European Conference on Speech Communication and Technology*.
- Massaro, D. W., & Palmer Jr, S. E. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Mit Press.
- Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience*, 26(3), 606–620.
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56(2), 154–178.
<https://doi.org/10.1002/dev.21177>
- McNeill, D. (1985). So You Think Gestures Are Nonverbal? *Psychological Review*, 92(3), 350–371. <https://doi.org/10.1037/0033-295X.92.3.350>
- Mehler, J., Sebastian-Gallés, N., & Nespors, M. (2004). Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. *The New Cognitive Neuroscience*, 3, 825–836.
- Mehler, J., Gervain, J., Macagno, F., Cogoi, S., & Pen, M. (2008). *The neonate brain detects speech structure*. 105(37).
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoni, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*.
[https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Meier, R. P., & Willerman, R. (1995). Prelinguistic gesture in deaf and hearing infants. In K. Emmorey & J. S. Reilly (Eds.), *Language, gesture, and space* (pp. 391–409). Lawrence Erlbaum Associates, Inc.
- Ménard, L., Leclerc, A., Brisebois, A., Aubin, J., & Brasseur, A. (2008). Production and perception of French vowels by blind and sighted speakers. *Proceedings of ISSP 2008 - 8th International Seminar on Speech Production*, (January), 197–200.

- Ménard, L., Toupin, C., Baum, S. R., Drouin, S., Aubin, J., & Tiede, M. (2013). Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, *134*(4), 2975–2987.
<https://doi.org/10.1121/1.4818740>
- Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*(1), 1–9.
<https://doi.org/10.1111/desc.12701>
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340.
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.
- Morgan, J. L., & Demuth, K. (1996). Signal to syntax: Bootstrapping from speech to grammar in early acquisition. In *Signal to Syntax*.
- Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In *Age and the rate of foreign language learning* (pp. 1–40).
- Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes : A Balancing Act across the Lifespan. *Trends in Neurosciences*, *39*(8), 567–579.
<https://doi.org/10.1016/j.tins.2016.05.003>
- Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: new evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 912.
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12.
- Nespor, M. A., Guasti, M. T., & Christophe, A. (1996). Selecting word order: the Rhythmic Activation Principle. *Interfaces in Phonology*, 1–26.

- Nespor, M., Shukla, M., Van De Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*. <https://doi.org/10.1418/28093>
- Özçalışkan, Ş., & Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Developmental Science*, *13*(5), 752–760.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9-B17.
- Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. John Benjamins Publishing.
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*. [https://doi.org/10.1016/S0163-6383\(99\)00003-X](https://doi.org/10.1016/S0163-6383(99)00003-X)
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*. <https://doi.org/10.1111/1467-7687.00271>
- Pennisi, A., & Falzone, A. (2016). *Darwinian biolinguistics*. Springer.
- Pennisi, A., & Falzone, A. (2020). *The Extended Theory of Cognitive Creativity* (S. I. Publishing., ed.).
- Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, *251*(5000), 1493-1496.
- Pinker, Steven & MacWhinney, B. (1987). The bootstrapping problem in language acquisition. In *Mechanisms of language acquisition* (pp. 399-441.).
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology*, *20*, 421–435.
- Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*. <https://doi.org/10.1177/0956797614568320>

- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Romaine, S. (1995). *Bilingualism*. Wiley-Blackwell.
- Rönnerberg, J., Samuelsson, S., Lyxell, B., Campbell, R., Dodd, B., & Burnham, D. (1998). Conceptual constraints in sentence-based lipreading in the hearing-impaired. In *Hearing by eye: II. The psychology of speechreading and auditory-visual speech* (pp. 143-153). Psychology Press.
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception and Psychophysics*. <https://doi.org/10.3758/BF03211902>
- Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, 33(12), 2329-2337.
- Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R., & Münte, T. F. (2018). Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain Imaging and Behavior*. <https://doi.org/10.1007/s11682-017-9694-y>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928. <https://doi.org/10.1126/science.274.5294.1926>
- Schaadt, G., van der Meer, E., Pannekamp, A., Oberecker, R., & Männel, C. (2019). Children with dyslexia show a reduced processing benefit from bimodal speech information compared to their typically developing peers. *Neuropsychologia*, 126(February 2017), 147-158. <https://doi.org/10.1016/j.neuropsychologia.2018.01.013>
- Searle, J. L. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). *A Bilingual Advantage in Visual Language Discrimination in Infancy*. <https://doi.org/10.1177/0956797612436817>
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232.

- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2). [https://doi.org/10.1016/S0010-0277\(99\)00047-5](https://doi.org/10.1016/S0010-0277(99)00047-5)
- Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, *69*(2), 218–231.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.
- Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.1907309>
- Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., & Watkins, K. E., Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain and Language*, *126*(1), 1–7.
- Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236.
- Tecumseh Fitch, W. (2010). The evolution of language. In *The Evolution of Language*. <https://doi.org/10.1017/CBO9780511817779>
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*. <https://doi.org/10.1016/j.cognition.2008.05.009>
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*. <https://doi.org/10.1186/1471-2202-10-21>
- Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7.1*, 53–71. https://doi.org/10.1207/s15327078in0701_5

- Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, 52(3), 314-320.
- Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin and Review*, 9(2), 335-340. <https://doi.org/10.3758/BF03196290>
- Turner, A. C., McIntosh, D. N., & Moody, E. J. (2015). Don't Listen With Your Mouth Full: The Role of Facial Motor Action in Visual Speech Perception. *Language and Speech*, 58(2), 267-278. <https://doi.org/10.1177/0023830914542305>
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*. <https://doi.org/10.1016/j.specom.2006.10.003>
- van Laarhoven, T., Keetels, M., Schakel, L., & Vroomen, J. (2018). Audio-visual speech in noise perception in dyslexia. *Developmental Science*. <https://doi.org/10.1111/desc.12504>
- Vouloumanos, A., & Werker, J. F. (2007). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science*, 10(2), 169-171.
- Vouloumanos, Athena, & Werker, J. F. (n.d.). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science*, 10(2), 169-171. <https://doi.org/10.1111/j.1467-7687.2007.00551.x>
- Vouloumanos, Athena, & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, 7(3), 270-276. <https://doi.org/10.1111/j.1467-7687.2004.00345.x>
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 252-258.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37(1), 159-170.

- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.
[https://doi.org/10.1016/S0028-3932\(02\)00316-0](https://doi.org/10.1016/S0028-3932(02)00316-0)
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*.
<https://doi.org/10.1126/science.1137686>
- Werker, Janet F. & Gervain, J. (2013). Speech perception in infancy: A Foundation for Language Acquisition. In *The Oxford handbook of developmental psychology* (Vol. 1, pp. 909–925). <https://doi.org/10.1126/science.171.3968.303>
- Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, 37(2), 278–286.
<https://doi.org/10.1037/h0080725>
- Werker, Janet F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63.
- Werker, Janet F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*.
[https://doi.org/10.1016/S0163-6383\(02\)00093-0](https://doi.org/10.1016/S0163-6383(02)00093-0)
- White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C., Wasim, N., ... Shergill, S. S. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Frontiers in Human Neuroscience*, 8(AUG), 1–12. <https://doi.org/10.3389/fnhum.2014.00565>
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.
- Wilson, Stephen M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1), 316–325. <https://doi.org/10.1016/j.neuroimage.2006.05.032>
- Worster, E., Pimperton, H., Ralph-Lewis, A., Monroy, L., Hulme, C., & MacSweeney, M. (2018). Eye Movements During Visual Speech Perception in Deaf and Hearing Children. *Language Learning*, 68(June), 159–179. <https://doi.org/10.1111/lang.12264>

Yeung, H. H., & Werker, J. F. (2013). Lip Movements Affect Infants' Audiovisual Speech Perception. *Psychological Science, 24*(5), 603–612.
<https://doi.org/10.1177/0956797612458802>

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science, 12*(5), 798-814.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*(10), 946-953.