DOI: 10.17469/O2111AISV000021

FEDERICA CAVICCHIO, MIRKO GRIMALDI

I training audio e audiovisivi migliorano in misura uguale la percezione segmentale della seconda lingua

Audio and audiovisual training equally enhance second language segmental perception

This study investigates the efficacy of high variability phonetic training, using audiovisual stimuli, in improving the perception of L2 Southern Standard British English (SSBE) phonemes by L1 Italian speakers. Participants underwent an audio pre-test and were assigned to audio, audiovisual, or control training groups. The post-test results revealed that audio and audiovisual training significantly improved the phoneme identification accuracy compared to the control group. However, during the training, the audiovisual information positively influenced the identification of L2 vowels. This improvement is, however, temporary. According to the *Analysis by Synthesis model*, our findings suggest that learners have the acoustic signal as privileged cue for building their phonetic-phonological representations. Their learning is facilitated only when the acoustic signal contains information related to distinctive features already active in the contrastive system of the L1. Further exploration of the interactions between auditory and articulatory information is essential for developing effective teaching approaches to L2 phonemes perception.

Keywords: high variability phonetic training, audiovisual speech, second language learning, second language sound perception.

1. *Introduzione*

L'*High Variability Phonetic Training* (HVPT) è uno dei metodi più utilizzati per migliorare le abilità di percepire e produrre i suoni di una seconda lingua straniera (L2). Sebbene il termine *High Variability Phonetic Training* sia stato coniato da Iverson, Hazan & Bannister (2005), le prime ricerche che utilizzano l'HVPT sono iniziate ben prima, stimolate dal dibattito sull'ipotesi del periodo critico di acquisizione della lingua e la sua validità in merito all'acquisizione della L2 (Flege, 1987; Major, 1987). Con l'accumularsi nel tempo di ricerche che usano l'HVPT, questo metodo si è confermato in grado di generare cambiamenti misurabili nella percezione e produzione dei fonemi L2 degli studenti adulti.

Nel primo studio in cui l'HVPT è stato utilizzato, Logan, Lively & Pisoni (1991) hanno addestrato sei apprendenti giapponesi a identificare il contrasto inglese /l/-/r/, assente nel loro sistema consonantico, usando un compito di identificazione seguito da feedback. Gli studenti sono stati esposti a 68 coppie minime di parole contenenti i suoni consonantici /l/-/r/in posizioni iniziale, mediale o finale di sillaba

pronunciate da cinque diversi parlanti. Dopo 40 minuti di training, gli studenti erano significativamente più abili a discriminare /l/-/r/e questa conoscenza si era generalizzata sia rispetto a parole nuove non addestrate sia rispetto ad un nuovo parlante che gli studenti non avevano sentito durante l'addestramento. L'accuratezza variava in base al contesto fonetico in cui si verificava il contrasto /l/-/r/. In particolare, gli studenti erano più bravi a identificare il contrasto quando si trovava in posizione finale di sillaba rispetto alla posizione iniziale. Logan *et al.* (1991) hanno anche scoperto che il contesto influenzava il tempo di reazione. Studi successivi hanno confermato i risultati di Logan *et al.*, 1991, generalizzandoli anche rispetto ai fonemi vocalici. Per quanto riguarda gli stimoli uditivi, nella maggior parte degli studi questi erano prodotti da 2 fino a 30 parlanti nativi. Mentre la norma è usare stimoli naturali per l'addestramento e i test di verifica, uno studio ha anche utilizzato la sintesi vocale per produrre gli stimoli L2 (Qian, Chukharev-Hudilainen & Levis, 2018).

In ricerche sulla comprensione lessicale, svariati studi hanno impiegato metodologie come il priming (per esempio, Broersma, 2012) e l'eye-tracking (ad esempio, Cutler, Weber & Otake, 2006; Escudero, Hayes-Herb & Mitterer, 2008; Weber e Cutler, 2004) per analizzare le differenze nell'elaborazione linguistica e lessicale tra parlanti L2 e parlanti nativi (L1). I risultati indicano che i parlanti L2, a differenza dei parlanti L1, tendono a mostrare tempi di reazione più lunghi per le parole che presentano contrasti fonetici tipici della L2, come ad esempio /æ/-/ʌ/. Invece, i parlanti L1 reagiscono più rapidamente a tali contrasti. Questo suggerisce che esiste una differenza significativa nel modo in cui i parlanti nativi e quelli L2 attivano lessicalmente le parole dei contrasti fonetici caratteristici di una determinata lingua. Il contrasto viene quindi percettivamente assimilato a un suono L1 (ad esempio, la vocale L1 /a/ per il contrasto /kæp/-/kʌp/). La difficoltà nel distinguere suoni L2 percettivamente simili aumenta la competizione lessicale, rendendo difficile il riconoscimento delle parole e potenzialmente portando alla confusione delle parole L2 in coppia minima. Neppure un'accurata categorizzazione dei suoni contrastivi in L2 sembra garantire un'accurata discriminazione a livello lessicale; infatti, le parole in coppia minima possono essere percepite come omofone se il contrasto target non è stato accuratamente codificato lessicalmente dai parlanti L2 (Darcy, Dekydtspotter, Sprouse, *et al.*, 2012). Di conseguenza, l'addestramento HVPT con parole in coppia minima può essere meno efficace rispetto all'uso di non-parole per le quali i parlanti L2 non hanno ancora sviluppato rappresentazioni lessicali. Thomson (2011) ha condotto una serie di training con HVPT modificando i contesti degli stimoli uditivi. Un gruppo di studenti L2 è stato addestrato a percepire vocali inglesi all'interno di non-parole CV prodotte da parlanti nativi, mentre un altro gruppo è stato addestrato con non-parole manipolate sinteticamente in modo tale che le vocali fossero raddoppiate in lunghezza. Entrambi i gruppi hanno dimostrato un miglioramento nell'identificazione delle vocali addestrate a prescindere dal tratto di lunghezza delle vocali.

1.1 Stimoli e tipi di training con l'HVPT

Nel HVPT gli stimoli uditivi vengono solitamente presentati in sessioni bloccate per parlante L1, come suggerito da Logan *et al.* (1991) e Lively, Logan & Pisoni (1993), invece di utilizzare più parlanti all'interno di una singola seduta di training. Tuttavia, studi più recenti hanno usato fino a 30 parlanti L1 per sessione (ad esempio, Thomson, 2016a; Thomson, Derwing, 2016).

L'uso del paradigma Forced Choice IDentification (FCID) sembra, in generale, essere più efficace della discriminazione AX, sia per le consonanti (Flege, 1995a) che per le vocali (Rato, Rauber, 2015; Carlet, 2017). Nel paradigma FCID, gli studenti devono identificare il fonema appena ascoltato, utilizzando simboli fonetici o ortografici (ad esempio, "hai sentito /l/ o /r/?" vs "hai sentito l o r?") dopo aver ascoltato uno stimolo uditivo. Resta invece ancora aperta la questione del training adattivo, ovvero dei cambiamenti della struttura del training in risposta alle prestazioni dei singoli studenti L2. Diversi studi hanno utilizzato una qualche forma di training HVPT adattivo (Pruit, Jenkings & Strange, 2006; Iverson, Evans, 2009; Lengeris, Hazan, 2010), ma nessuno di questi ha paragonato l'effetto dell'apprendimento adattivo con quello non adattivo.

La durata del training solitamente varia da poche ore ad alcune settimane, con risultati omogenei (Logan *et al.*, 1991; Thomson, 2016b). Il numero di sessioni di formazione e la loro durata cumulativa variano notevolmente da studio a studio. Ad esempio, Cebrian, Carlet (2014) hanno completato quattro sessioni di training, per un totale di tre ore, mentre i partecipanti di Nishi, Kewley-Port (2007, 2008) hanno completato nove sessioni, per un totale di 13,5 ore. Non sembra esistere una corrispondenza tra la durata del training e il numero di fonemi utilizzati per il training. Ad esempio, Logan *et al.* (1991) ha mirato solo al contrasto /l/-/r/ e Thomson (2016a) ha sviluppato un training per dieci vocali: in entrambi gli studi gli apprendenti sono stati sottoposti a dieci ore di training in totale senza un evidente vantaggio della durata dei training sul livello di apprendimento dei fonemi.

1.2 Generalizzazione e mantenimento nel tempo dell'apprendimento percettivo

Uno dei vantaggi più importanti dell'HVPT è che l'apprendimento percettivo dei suoni L2 oggetto del training si generalizza rispetto a nuove istanze dello stesso suono e l'apprendimento viene mantenuto nel tempo. L'effetto sull'apprendimento percettivo L2 è solitamente generalizzato anche rispetto a nuovi parlanti che hanno prodotto le stesse non-parole o sillabe usate durante il training (Lively *et al.*, 1993). Tuttavia, la generalizzazione potrebbe non avvenire per tutti i suoni oggetto del training, come ad esempio nell'inglese /ð/ (Cebrian, Carlet, 2014).

Ci sono risultati contrastanti riguardo ai test di generalizzazione dell'HVPT rispetto a nuove non-parole. Alcuni studi hanno mostrato che il training con parlanti multipli ha portato a discriminare anche nuove non-parole prodotte da nuovi parlanti (Lively *et al.*, 1993), mentre il training con un solo parlante non sembra portare alla generalizzazione di non-parole non familiari prodotte dallo stesso parlante. Altri studi hanno trovato solo un trasferimento parziale dell'apprendimento percettivo agli stessi suoni in nuovi

contesti. Iverson *et al.* (2005) hanno riportato che la percezione del contrasto inglese /l/ e /r/ da parte di parlanti L2 giapponese si estende a nuove parole, principalmente a quelle che iniziano con /l/ e /r/. L'apprendimento non era invece generalizzato quando le stesse consonanti si trovano in posizione mediale tra vocali. Thomson (2011) ha evidenziato che, in produzione, 10 vocali inglesi nei contesti /bV/ e /pV/ venivano generalizzate, dopo il training HVPT, anche in contesti /zV/ e /sV/, ma non in contesti /gV/ e /kV/. Tuttavia, in percezione è stato riscontrato il comportamento opposto. Le vocali del training in contesti /bV/ e /pV/ erano generalizzate nei contesti /gV/ e /kV/, ma non nei contesti /zV/ e /sV/ (Thomson, 2012). Altre ricerche invece non hanno trovato alcun trasferimento a nuovi contesti. Tajima *et al.* (2008) hanno riportato che, dopo il training, i parlanti anglofoni sono in grado di percepire la lunghezza dei fonemi del giapponese e di generalizzarla a nuovi parlanti, ma non hanno trovato un trasferimento della percezione vocalica a nuovi contesti.

Un'altra domanda rilevante nell'utilizzo dell'HVPT è fino a che punto sia possibile trasferire il training fonetico fatto in contesto di parole isolate a parole che si trovano nel contesto di un discorso. Hirata (2004) ha addestrato studenti inglesi a percepire i contrasti di lunghezza delle vocali e delle consonanti giapponesi. Un gruppo ha ascoltato parole pronunciate in isolamento, mentre un altro gruppo ha ascoltato le stesse parole pronunciate in una frase. I risultati dimostrano che entrambe le condizioni hanno portato a un significativo miglioramento. Inoltre, il trasferimento dall'apprendimento in contesti di frasi a quello in parole isolate era maggiore del trasferimento dall'addestramento di parole isolate al contesto in frasi.

L'apprendimento fonetico dovrebbe non solo essere generalizzabile, ma anche mantenuto nel lungo termine. Diversi studi hanno dimostrato come, dopo il training con HVPT, l'apprendimento fonetico è mantenuto dopo due settimane (Flege, 1995b), un mese (Thomson, 2012), due mesi (Rato, Rauber, 2015), tre mesi (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Wang, Munro, 2004; Nishi, Kewley-Port, 2007; Carlet, 2017) e quattro mesi (Iverson, Evans, 2009). Solo due studi hanno indagato la persistenza dell'apprendimento oltre i quattro mesi. Lively, Pisoni, Yamada *et al.*, (1994) hanno scoperto che gli studenti giapponesi miglioravano la loro pronuncia del contrasto inglese /l/-/r/ fino a tre mesi dopo il training, ma dopo sei mesi si presentava un declino e i punteggi di identificazione non erano significativamente diversi dai punteggi pre-test. Viceversa, Wang, Spence, Jongman & Sereno (1999) hanno trovato che l'apprendimento percettivo può persistere senza declino per almeno sei mesi.

Thomson, Derwing (2016) hanno studiato se il training con non-parole portasse a un miglioramento nella produzione di 10 vocali inglesi L2 rispetto all'addestramento con parole L2 contenenti le medesime vocali. I 31 partecipanti (11 L1 cinesi, 5 L1 spagnoli, 3 L1 etiopi parlanti amarico, 3 L1 arabi, 2 L1 russi, 2 L1 francesi, 2 L1 Punjabi, 1 L1 rumeno, 1 L1 zo'é e 1 L1 igbo) sono stati divisi in due gruppi. Il primo gruppo è stato addestrato a percepire le dieci vocali inglesi con non-parole contenenti le vocali in contesto CVC. Questo gruppo ha significativamente migliorato la pronuncia di 70 parole reali contenenti le stesse vocali addestrate. Al contrario, il secondo gruppo addestrato a percepire le stesse vocali, utilizzando però le 70 parole reali, non aveva

migliorato significativamente la pronuncia delle parole stesse. Questo risultato è coerente con quanto riportato da Nishi, Kewley-Port (2007) e Carlet (2017), che nei loro studi hanno usato non-parole per addestrare la percezione fonetica delle vocali dell'inglese L2. Entrambi gli studi riportano che l'addestramento con non-parole veniva generalizzato a parole che contenevano le vocali addestrate.

1.3 HVPT con stimoli audiovisivi

È noto che, negli adulti, le aree senso-motorie e uditive sono attivate congiuntamente durante l'acquisizione dei suoni della L2 (Kuhl, Ramírez, Bosseler *et al.*, 2014). Hazan, Sennema, Iba & Faulkner (2005) hanno condotto uno studio per verificare se l'aggiunta di stimoli visivi nell'HVPT potesse portare a un miglioramento del riconoscimento fonetico rispetto all'uso di stimoli solo uditivi. Hazan *et al.* (2005) hanno usato stimoli audiovisivi che mostravano i movimenti delle labbra dei parlanti mentre producevano una serie di non-parole con i contrasti /v/-/b/ e /l/-/r/. I risultati dello studio hanno mostrato che gli stimoli audiovisivi portavano a un miglioramento significativo dell'identificazione delle consonanti nella coppia minima /b/-/v/, ma non nella coppia minima /l/-/r/. Questo risultato misto potrebbe essere dovuto alla natura più evidente dal punto di vista visivo dei suoni labiodentali e bilabiali della coppia /b/-/v/. Hazan, Sennema, Faulkner *et al.*, (2006) hanno indagato l'impatto della componente visiva sulla percezione del parlato da parte di parlanti L1 e L2 in ascolto degradato. Lo studio presentava frasi degradate con rumore, riverbero e compressione dell'ampiezza, con o senza l'input visivo corrispondente dei movimenti labiali di chi palava. I risultati hanno mostrato che sia i parlanti L1 madrelingua che L2 beneficiavano dei segnali visivi. Tuttavia, il beneficio variava a seconda del tipo specifico di degrado del suono e del *background* linguistico dei partecipanti (L1 o L2). I risultati suggeriscono che i segnali visivi possono svolgere un ruolo importante nell'aiutare la percezione del parlato in condizioni di ascolto difficili sia per i parlanti L1 che per quelli L2. Similmente, Hannah, Wang, Jongman, *et al.*, (2017) mostrano che in un compito di identificazione di sillabe in mandarino sia i parlanti L1 che L2 beneficiano delle informazioni visive.

Uno dei possibili motivi di questi risultati è che l'input visivo fornisce ulteriori segnali sui movimenti articolatori associati a specifici fonemi. Gli input visivi potrebbero quindi aiutare gli studenti L2 adulti a sviluppare una rappresentazione più accurata e sfumata del suono L2. Osservando un parlante madrelingua produrre un particolare suono, gli studenti L2 possono osservare i movimenti delle labbra, della lingua e della mascella necessari per produrre quel suono. Queste informazioni visive potrebbero aiutare gli studenti a sviluppare una rappresentazione più dettagliata e accurata del suono, facilitandone il successivo riconoscimento e produzione. Inoltre, alcuni fonemi L2 possono essere difficili da distinguere basandosi solo su segnali uditivi, ma possono essere distinti più facilmente se accompagnati da segnali visivamente salienti. Nel complesso, incorporare l'input visivo nell'apprendimento dei fonemi L2 potrebbe quindi fornire ulteriori segnali e informazioni che supportano la percezione e la produzione dei suoni della L2.

1.4 Lo studio presente

Lo studio di Hazan *et al.* (2005) ha dimostrato che l'uso di stimoli audiovisivi può migliorare l'identificazione dei suoni L2 in un paradigma HVPT. Tuttavia, non è ancora chiaro quale sia il ruolo delle specifiche informazioni articolatorio-visive nell'apprendimento fonetico della L2. Le informazioni articolatorio-visive, fornendo un supporto complementare e integrativo all'apprendimento fonologico uditivo, potrebbero infatti aiutare i parlanti L2 a percepire meglio le caratteristiche distintive dei fonemi della L2. L'apprendimento dei movimenti articolatori specifici dei fonemi della L2 può inoltre sviluppare una maggiore consapevolezza delle differenze fonologiche tra i fonemi L2 e quelli della L1. Inoltre, gli studenti devono consolidare le categorie fonologiche L2 apprese e generalizzarle in contesi diversi. In situazioni linguistiche immersive, questo avviene attraverso l'uso continuo della lingua e l'esposizione a una varietà di contesti linguistici. In questo senso, l'uso dell'HVPT, accoppiato all'uso delle informazioni articolatorie e uditive, potrebbe aiutare a generalizzare e consolidare le distinzioni fonologiche L2 oggetto di training. Per aumentare l'efficacia del training HVPT con stimoli audiovisivi, potrebbe essere utile utilizzare stimoli visivamente salienti come, ad esempio, l'apertura maggiore o minore della bocca o la presenza o meno dell'arrotondamento delle labbra. Per testare l'ipotesi che *features* visivamente salienti aiutino a consolidare le categorie fonologiche della L2, abbiamo utilizzato le informazioni articolatorie visive di due contrasti vocalici del Southern Standard British English (SSBE), /e-æ/ e /ʌ-ɒ/. Questi contrasti sono particolarmente difficili da discriminare uditivamente per i parlanti di italiano salentino (Escudero, Sisinni & Grimaldi, 2014), che è caratterizzato da cinque fonemi vocalici: /i, ɛ, a, ɔ, u/. La scelta di coppie minime specifiche, già testate uditivamente in altri studi sui parlanti L1 dell'italiano salentino, ci permette di verificare se l'aggiunta di informazioni articolatorie visive supporta lo sviluppo di rappresentazioni mnemoniche inerenti alle proprietà acustico-articolatorie dei contrasti fonologici L2. Secondo la teoria dei *Tratti Distintivi*, è ben noto che la coppia minima /e-æ/ basa la sua contrastività fonologica sul tratto acustico-articolatorio [±basso] (che, secondo il modello teorico, ha anche un ruolo nel processo percettivo): /e/ è caratterizzato dai tratti [-alto, -basso, -posteriore, -arrotondato], mentre /æ/ dai tratti [-alto, +basso, -posteriore, -arrotondato]. L'opposizione fonologica /ʌ-ɒ/ si fonda, invece, sui tratti [±basso] e [±arrotondato]: /ʌ/ è un fonema [-alto, -basso, +posteriore, -arrotondato], mentre /ɒ/ è [-alto, +basso, +posteriore, +arrotondato] (cfr. Odden, 2005).

In linea con il modello *Analysis by Synthesis* (Halle, 2001), il cervello di chi apprende un sistema linguistico è in grado di estrarre rappresentazioni audio-motorie dall'analisi del segnale acustico (integrando unità percettive e unità motorie); in questo modo, viene generata una mappa audio-motoria funzionale alla scoperta dei principi che oppongono i suoni linguistici fra loro in termini di tratti distintivi. Al contrario, secondo la *Motor Theory of Speech Perception* (Liberman, Mattingly, 1985) l'apprendimento di un sistema fonologico avviene

attraverso l'analisi e la memorizzazione dei gesti articolatori direttamente implicati nella generazione del segnale acustico. A generare le rappresentazioni mentali dei contrasti fonologici sarebbero, quindi, le configurazioni articolatorie astratte coinvolte nella produzione dei suoni percepiti: in questo modello, dunque, gli atomi dei suoni linguistici non sono i tratti distintivi, ma proprio le configurazioni articolatorie astratte.

L'apprendimento di contrasti fonologici della L2 in età adulta diventa un banco di prova per le due teorie. Con questo lavoro, infatti, ci proponiamo di verificare:

– Se l'utilizzo di training HVPT con informazioni articolatorie audiovisive chiaramente visibili migliora l'apprendimento percettivo dei suoni L2 da parte di parlanti L1 dell'italiano salentino rispetto al training solo uditivo.

In parallelo, il nostro obiettivo è anche testare gli assunti dei due modelli appena discussi all'interno dei processi apprendimento fonetico-fonologico di una L2 (limitatamente allo sviluppo percettivo), e quindi verificare:

– Se anche gli apprendenti adulti, come i bambini, fanno principalmente ricorso al segnale acustico per costruirsi rappresentazioni fonetico-fonologiche dei suoni L2, le informazioni visive non dovrebbero produrre nessun vantaggio percettivo. Al contrario, se, come ipotizza la *Motor Theory of Speech Perception*, sono i gesti articolatori a stimolare la generazione di rappresentazioni fonologiche, le informazioni audio-visive aggiunte dovrebbero facilitare la categorizzazione di quei fonemi L2 caratterizzati da specificità articolatorie come l'arrotondamento delle labbra.

2. *Metodo*

2.1 Stimoli

Abbiamo video-registrato 40 non-parole (20 per la coppia minima /e-æ/ e 20 per la coppia minima /ʌ-ɒ/) prodotte da sei madrelingua di SSBE (tre maschi e tre femmine) di età compresa tra 35 e 56 anni. Seguendo studi precedenti (e.g. Carlet, 2017), le vocali sono state prodotte in contesto CVC assieme a consonanti bilabiali occlusive e nasali (/p, b, m/), velari occlusive (/k, g/), labiodentali fricative (/f, v/), e dentali occlusive (/t, d/). Per elicitare la produzione delle non-parole contenenti i fonemi selezionato per lo studio, i partecipanti hanno pronunciato le non-parole all'interno di frasi del tipo "*it rhymes with bed* [pausa] *tet*". L'audio e il video sono stati registrati con l'ausilio di un microfono Speedlink collegato esternamente a una telecamera Sony HDR. La telecamera riprendeva la parte inferiore del viso del parlante, concentrandosi sui movimenti articolatori delle labbra e della lingua (vedasi Fig. 1, pannello Audiovisivo). Due parlanti madrelingua SSBE hanno giudicato su una scala da 1 a 5 le produzioni audiovideo, valutando la conformità delle non-parole con i fonemi del SSBE (scala da 1 non conforme - 5 totalmente conforme). Le non-parole che hanno ricevuto un grado di giudizio pari o superiore

a quattro sono state selezionate per il training. I videoclip e i file audio delle non-parole sono state estratte dalle frasi con il software VLC.

Per il gruppo di controllo sono stati usati degli esercizi di comprensione del testo seguiti da relative domande a scelta multipla. Gli esercizi sono stati realizzati per studenti di inglese L2 di livello CEFR B1 e sono stati tratti dal sito web https://test-english.com/reading/b1/.

2.2 Partecipanti e gruppi di training

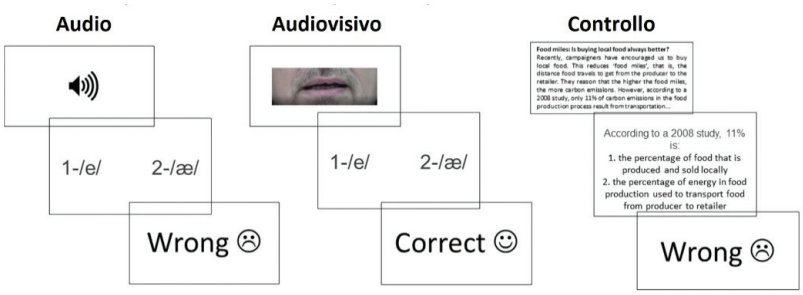
Allo studio hanno preso parte 66 parlanti di italiano salentino (età media dei partecipanti=19,5, M= 20, F= 46, livello CEFR inglese B1=51, A2=15). Tutti i partecipanti allo studio erano studenti del primo anno del corso di Lingue, culture e letterature straniere dell'Università del Salento e hanno ricevuto istruzioni riguardanti la fonetica acustica e articolatoria durante il corso di Linguistica generale che comprendevano anche le vocali prese in oggetto da questo studio. I parlanti selezionati erano tutti residenti in Salento da almeno 15 anni. Nessuno di loro era bilingue, e tutti avevano studiato inglese solo in contesto scolastico per almeno 8 anni, con soggiorni in paesi anglofoni non superiori ai 60 giorni. I partecipanti sono stati assegnati in modo casuale a tre gruppi sperimentali: training audio, audiovisivo o gruppo di controllo. I tre gruppi sono stati bilanciati per genere e livello di inglese.

2.3 Procedimento

Tutti i partecipanti hanno eseguito un pre-test uditivo di identificazione di 80 stimoli audio (40 non-parole x 2 parlanti nativi, 1 maschio e 1 femmina). Gli stimoli sono stati divisi in due sessioni di coppie minime (/e-æ/e /ʌ-ɒ/) in maniera randomizzata. Le sessioni sono state presentate ai partecipanti in modo controbilanciato (metà partecipanti hanno identificato prima i suoni della coppia minima /e-æ/e metà quelli della coppia minima /ʌ-ɒ/). Alla fine del pre-test i partecipanti hanno ricevuto un *feedback* globale sulla loro performance. I partecipanti assegnati ai gruppi di training audio e audiovisivo hanno preso parte a quattro sessioni di training ad intervalli settimanali, durante i quali hanno identificato 80 non-parole prodotte da 2 diversi parlanti (di sesso maschile e femminile) ricevendo *feedback* immediati (giusto/sbagliato) dopo ogni stimolo audio o audiovisivo identificato. Le non-parole sono state presentate ai partecipanti in due blocchi controbilanciati (un blocco per /e-æ/e un blocco per /ʌ-ɒ/). All'interno dei blocchi gli stimoli erano randomizzati. I partecipanti del gruppo di controllo hanno invece sostenuto a intervalli settimanali un test di comprensione del testo seguito da una serie di risposte a scelta multipla e feedback (Fig. 1). Gli stimoli erano presentati in modo randomizzato. Tutti i partecipanti hanno gestito il training in maniera autonoma, senza costrizioni temporali, ma generalmente la durata media di ogni sessione di training era di 25 minuti. Una settimana dopo la fine del training tutti i partecipanti hanno fatto un post-test uditivo di identificazione, con 40 non-parole, registrate da quattro

parlanti madrelingua, due (un maschio e una femmina) già ascoltati in precedenza durante il training e due nuovi parlanti (un maschio e una femmina) mai ascoltati in precedenza, e 16 nuove non-parole pronunciate sia dai due parlanti del training che dai due nuovi parlanti. Nel nostro studio, abbiamo addestrato i partecipanti utilizzando non-parole. Di conseguenza, per valutare la generalizzazione delle capacità percettive a nuovi suoni non inclusi nell'addestramento, abbiamo utilizzato nel post-test nuove non-parole.

Figura 1 - *Procedimento dei training audio, audiovisivo e del gruppo di controllo*



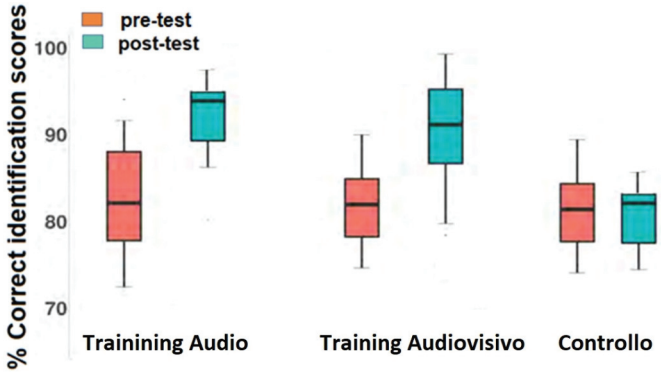
2.4 Analisi statistiche

I dati raccolti sono stati analizzati con una regressione logistica ad effetti misti, nella quale l'accuratezza dell'identificazione delle vocali era la variabile dipendente (1= vocale identificata, 0= vocale non identificata) e con il test (pre-test vs post-test), il metodo di training (audiovisivo, audio o gruppo di controllo) e le vocali /e/, /æ/, /ʌ/ e /ɒ/) come variabili indipendenti. Sono stati inoltre introdotti effetti casuali per partecipanti e item (non-parole). La regressione logistica è stata condotta con il software R versione 3.6 (R Core Team, 2022).

3. *Risultati*

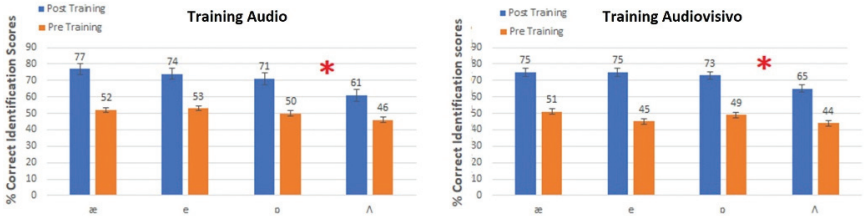
I risultati del modello statistico hanno mostrato una differenza significativa tra il pre-test e il post-test nell'identificazione delle vocali SSBE per i partecipanti al training audio rispetto ai partecipanti del gruppo di controllo (Est. = 0,77, SE=0,1, p<0,001) e tra il gruppo del training audiovisivo e il gruppo di controllo (Est. =0,97, SE=0,1, p<0,001) ma non tra i gruppi di training audiovisivo e audio (Est. = 0,2, SE=0,1, p=0,87, Fig. 2). In linea con risultati sul paradigma HVPT precedenti, non è stata trovata una differenza significativa né tra l'identificazione di nuove non-parole nel post-test e di non-parole usate durante il training (Est. =0.09, S. E=0.08, p=0.25) né tra l'identificazione di non-parole pronunciate dal nuovo parlante e dai parlanti ascoltati precedentemente durante il training (Est=0.04, S.E.=0.1, p=0.66).

Figura 2 - *Risultati del modello statistico: c'è una differenza significativa tra i risultati del pre-test e post-test nei partecipanti del gruppo audio e audiovisivo, ma non tra i partecipanti del gruppo di controllo. Inoltre, non ci sono differenze significative nei risultati del post-test tra i partecipanti del training audio e audiovisivo*



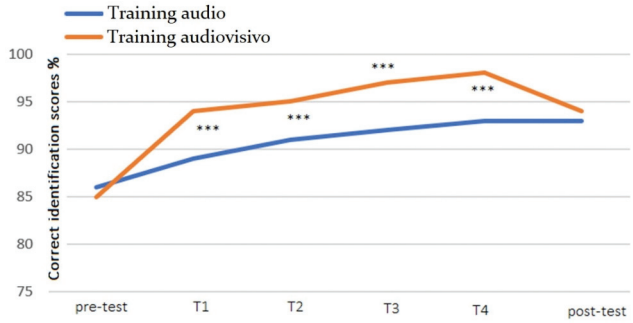
Esiste una differenza significativa tra l'accuratezza percettiva di /ʌ/ vs./ɒ/ sia per il gruppo del training audiovisivo (Est. = 0,96, SE=0,14, p<0,001) che per il gruppo che ha seguito il training audio (Est. =0,9, SE=0,14, p <0,001), suggerendo che per chi parla italiano salentino il contrasto vocalico SSBE più difficile da percepire è /ʌ/-/ɒ/ (Fig. 3). In particolare, la vocale /ʌ/ è quella che ha raggiunto i punteggi di identificazione più bassi tra tutte le vocali sia prima che dopo l'addestramento uditivo e audiovisivo.

Figura 3 - *Risultati del modello statistico: confrontando i risultati pre-training e post-training per i fonemi addestrati, si nota che il fonema /ʌ/ è quello significativamente più difficile da identificare (asterisco rosso)*



Un'ulteriore analisi statistica ha controllato se ci fosse una differenza significativa nell'identificazione dei fonemi del SSBE tra i partecipanti al training uditivo e audiovisivo. I risultati hanno dimostrato che i partecipanti del training audiovisivo identificano meglio i fonemi target rispetto ai partecipanti al traning solamente uditivo (Fig. 4, primo training (T1): Est=0.3, S.E.=0.1, p<0.001; secondo training (T2): Est=0.4, S.E.=0.1, p<0.001; terzo training (T3): Est=0.2, S.E.=0.1, p<0.001; quarto training (T4): Est=0.3, S.E.=0.1, p<0.001).

Figura 4 - *Confronto tra i risultati del pre-test (solo audio), training audio e audiovisivo a intervalli settimanali (T1= primo training, T2=secondo training, T3= terzo training, T4=quarto training) e post-test (solo audio). La percentuale di identificazione dei fonemi target è significativamente maggiore durante il training audiovisivo che in quello audio*



4. *Discussione e conclusioni*

I risultati dell'esperimento forniscono informazioni rilevanti rispetto alle ipotesi sperimentali elaborate. Con la prima ipotesi ci chiedevamo se l'utilizzo di training HVPT con l'aggiunta di informazioni articolatorie visive fosse più efficace nel migliorare l'identificazione dei suoni L2 del SSBE rispetto al solo training audio. In parallelo, rispetto a un processo di apprendimento dei suoni L2 in età adulta, ci siamo proposti di verificare alcuni assunti di base di due classici modelli teorici: l'*Analysis by Synthesis* e la *Motor Theory of Speech Perception* (cfr. 1.4).

I risultati hanno dimostrato che entrambi i gruppi sottoposti a training, sia quello audiovisivo che quello audio, hanno mostrato un miglioramento significativo nell'identificazione delle vocali del SSBE rispetto al gruppo di controllo. Tuttavia, non è emersa una differenza significativa tra i due gruppi di training audiovisivo e audio nel post-test. Questo suggerisce che l'aggiunta di informazioni articolatorie audiovisive non porta a un miglioramento nell'identificazione uditiva dei suoni L2 rispetto al solo training audio. Ciò sembrerebbe confermare gli assunti alla base del modello *Analysis by Synthesis*, secondo cui il parlante, anche adulto a questo punto, fa ricorso al segnale audio per generare rappresentazioni audio-motorie fonologicamente contrastive.

In particolare, i risultati del contrasto vocalico L2 /ʌ/-/ɒ/ dimostrano che – nonostante l'aggiunta delle informazioni articolatorie relative all'arrotondamento o meno delle labbra – i partecipanti hanno una percentuale di identificazione significativamente minore rispetto al contrasto /e-æ/, sia nel gruppo di training audiovisivo che nel gruppo di training audio. Inoltre, i bassi punteggi di identificazione della vocale /ʌ/ suggeriscono che i parlanti di italiano salentino hanno maggiori difficoltà nell'identificare questa vocale del SSBE, che viene confusa sistematicamente con la vocale /ɒ/.

Questo risultato richiede di essere adeguatamente interpretato, soprattutto in relazione ai modelli teorici precedentemente discussi. Nell'italiano salentino, che come abbiamo visto è caratterizzato da cinque fonemi vocalici, l'arrotondamento delle labbra – a differenza dei tratti [±alto] e [±posteriore] – non è un tratto distintivo che contribuisce alla contrastività fonematica. Infatti, il tratto [±arrotondato] in questo sistema è ridondante, in quanto le vocali /ɔ, u/, rispetto ad /i, ɛ/, sono entrambe [+posteriori] e [+arrotondate] e la vocale /a/ è caratterizzata solo dal tratto [+basso]: perciò il tratto [-arrotondato] di /a/ è predicibile. Dunque, il fatto che gli apprendenti adulti facciano ricorso al segnale acustico come risorsa privilegiata per generare rappresentazioni percettivamente contrastive permette loro di sviluppare una discriminazione fonemica solida per il tatto di altezza, come nel caso di /e-æ/, sfruttando un tratto già contrastivo nel sistema della L1; al contrario, lo stimolo acustico non sembra sufficiente per costruire una rappresentazione contrastiva robusta per il tratto [±arrotondato]. In altri termini, la risorsa articolatoria aggiunta nel training può agevolare il processo di identificazione, ma non sembra essere una informazione così saliente da riorientare la plasticità uditiva verso un contrasto fonologico L2 marcato da un tratto che nel sistema della L1 è ridondante.

Questa interpretazione ci pare supportata dai risultati riportati in Fig. 4. Da un lato, le informazioni audiovisive aggiunte producono un miglioramento del processo di discriminazione; dall'altro, le stesse informazioni (arrotondamento/non arrotondamento delle labbra) non sembrano sufficienti per generalizzare rappresentazioni fonemiche non presenti nel sistema contrastivo della L1. I nostri risultati ci sembrano in linea con quelli ottenuti da Hazan e colleghi (2005; 2006), i quali hanno evidenziato un miglioramento nelle capacità di discriminazione quando vengono fornite informazioni articolatorie audiovisive aggiuntive. Infatti, l'aggiunta delle informazioni articolatorie durante il training ha fornito ai partecipanti un vantaggio significativo nel riconoscimento dei suoni L2 rispetto al solo training audio. Potremmo dunque interpretare i nostri risultati come un supporto per l'ipotesi che l'attivazione dell'informazione motoria delle labbra giochi un ruolo nella percezione e nell'identificazione dei suoni del linguaggio, almeno durante la percezione audiovisiva. Questo potrebbe essere spiegato dal fatto che l'input visivo, che mostra i movimenti delle labbra dei parlanti, fornisce indizi motori aggiuntivi agli apprendenti. Il ricorso a questi indizi motori potrebbe quindi facilitare la corrispondenza tra l'input uditivo e la rappresentazione motoria del suono, migliorando così la percezione e l'identificazione dei suoni L2, ma non a tal punto, come nel nostro caso, da produrre la costruzione solida di rappresentazioni fonologiche distanti dalla L1. Infatti, quando nel post-test gli stimoli vengono presentati nella sola modalità audio emerge una netta discrepanza nella capacità di discriminare coppie di fonemi L2 differenziati rispetto ai tratti distintivi che hanno funzione contrastiva nel sistema della L1 ([±alto] vs. [±arrotondato]).

In sintesi, i risultati di questo studio indicano che sia il training audiovisivo che quello audio producono un miglioramento significativo nell'identificazione dei suoni L2 del SSBE per i parlanti L1 di italiano salentino. Il training audiovisivo,

dunque, non offre un vantaggio diretto nel rimodellare la plasticità uditiva degli apprendenti L2. E questo può dipendere dal fatto che, come abbiamo visto, i parlanti estraggono i parametri distintivi di un sistema fonologico direttamente dal segnale acustico (come assunto nel modello *Analysis by Synthesis*). Quando nel segnale acustico della L2 sono presenti informazioni relative a tratti distintivi contrastivamente già attivi nel sistema L1, queste vengono riutilizzate dagli apprendenti per costruire nuove rappresentazioni mnemoniche per i fonemi L2. Quando, invece, nel segnale acustico della L2 sono presenti informazioni relative a un tratto distintivo che nel sistema L1 non ha funzione contrastiva, gli apprendenti hanno difficoltà a costruirsi nuove rappresentazioni fonemiche stabili. Le ricerche future dovranno testare se in questi casi sia necessario un periodo di training più prolungato o se, in parallelo, associare al training uditivo un training articolatorio (aggiungendo, magari, un periodo di riflessione metalinguistica sulle proprietà fonetico-fonologiche dei suoni L2 che si stanno apprendendo). Infatti, se il bambino è in grado di scoprire in modo naturale le proprietà di qualsiasi sistema fonologico a cui è esposto in contesto naturale, l'adulto – come dimostrato da Kuhl e colleghi (2014) – mostra una differente attivazione del network neurale per avvicinarsi agli stessi obiettivi di apprendimento.

Nel complesso, questi risultati hanno importanti implicazioni per lo sviluppo di metodi didattici destinati agli adulti che apprendono una L2 in contesto formale. Nella progettazione e sperimentazione di training mirati (da utilizzare in futuro nella pratica didattica in classe) bisognerà, quindi, tenere nella giusta considerazione le differenze fonematiche fra il sistema L1 ed il sistema L2 e, in parallelo, pensare di sviluppare training flessibili in grado di adattarsi alle diverse proprietà della L2 rispetto alla L1 di partenza.

Bibliografia

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., Tohkura, Y.I. (1997). Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production. In *The Journal of the Acoustical Society of America*, 101(4), 2299-2310. https://doi.org/10.1121/1.418276

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. In *Language and Cognitive Processes*, 27(7-8), 1205-1224. http://dx.doi.org/10.1080/01690965.2012.660170

Carlet, A. (2017). L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study. PhD. Dissertation, Universitat Autonoma de Barcelona.

Cebrian, J., Carlet, A. (2014). Second-language learners' identification of target-language phonemes: A short-term phonetic training study. In *Canadian Modern Language Review*, 70(4), 474-499. https://doi.org/10.3138/cmlr.2318

Cutler, A., Weber, A., Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. In *Journal of Phonetics*, 34(2), 269-284. https://doi.org/10.1016/j.wocn.2005.06.002

Darcy, I., Dekydtspotter, L., Sprouse, R.A., Glover, J., Kaden, C., McGuire, M., Scott, J.H.G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition. In*Second Language Research*, 28, 5-40. http://dx.doi.org/10.1177/0267658311423455

Escudero, P., Hayes-Harb, R., Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. In *Journal of Phonetics*, 36, 345–360 https://doi.org/10.1016/j.wocn.2007.11.002

Escudero, P., Sisinni, B., Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. In *Journal of Acoustic Society of America*,135(3), 1577-84. https://doi.org/10.1121/1.4864477

Flege, J.E. (1987). A critical period for learning to pronounce foreign languages? In *Applied Linguistics*, 8(2), 162-177.

Flege, J. (1995a). Second language speech learning: Theory, findings, and problems. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross language research*. Timonium, MD: York Press, 233-277.

Flege, J.E. (1995b). Two procedures for training a novel second language phonetic contrast. In *Applied Psycholinguistics*, 16, 425-442. https://doi.org/10.1017/S0142716400066029

Halle, M. (2001). *Form Memory to Speech and Back*. Berlin-New York: De Gruyter.

Hazan, V., Sennema, A., Iba, M., Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. In *Speech Communication*, 47(3), 360-378. https://doi.org/10.1016/j.specom.2005.04.007

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. In *Journal of Acoustoc Society of America*, 119(3), 1740–1751. https://doi.org/10.1121/1.2166611

Hannah, B., Wang, Y., Jongman, A., Sereno, J.A., Cao, J., Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. In *Frontiers in Psychology*, 8(2051). https://doi.org/10.3389/fpsyg.2017.02051

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. In *The Journal of the Acoustical Society of America*, 116(4), 2384-2394. https://doi.org/10.1121/1.1783351

Houde, J.F., Jordan, M.I. (1998). Sensorimotor adaptation in speech production. In *Science*, 279, 1213-1216. https://doi.org/10.1126/science.279.5354.1213

Iverson, P., Hazan, V., Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. In *The Journal of the Acoustical Society of America*, 118(5), 3267-3278. https://doi.org/10.1121/1.2062307

Iverson, P., Evans, G.B. (2009). Learning English vowels with different first-language vowel systems II: Auditor/ training for native Spanish and German speakers. In *Journal of Acoustical Society of America*, 126(8), 866-877. https://doi.org/10.1121/1.3148196

Kuhl, P.K., Ramírez, R.R., Bosseler, A., Lin J.L., Imada, T. (2014). Infants' brain responses to speech suggest Analysis by Synthesis. Proceedings of the National Academy of Sciences. 111(31), 11238-11245. https://doi.org/10.1073/pnas.1410963111

Lengeris, A., Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. In *The Journal of the Acoustical Society of America*, 128(6), 3757-3768. https://doi.org/10.1121/1.3506351

Liberman, A.M., Mattingly, I.G. (1985). The motor theory of speech perception revised. In *Cognition*, 21(1), 1–36.

Lively, S.E., Logan, J.S., Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/, II: The role of phonetic environment and talker variability in learning new perceptual categories. In *The Journal of the Acoustical Society of America*, 94(3), 1242-1255.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. In *Journal of the Acoustic Societyof America*, 96(4), 2076-2087. https://doi: 10.1121/1.410149

Logan, J.S., Lively, S.E., Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/and/1/: A first report. In*The Journal of the Acoustical Society of America*, 89(2), 874-886.https://doi.org/10.1121/1.1894649

Major, R.C. (1987). Foreign accent: Recent research and theory. In*International Review of Applied Linguistics in Language Teaching*, 25(4), 185-202. https://doi.org/10.1515/iral.1987.25.1-4.185

Nishi, K., Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. In *Journal of Speech, Language, and Hearing Research*, 50(6), 1496-1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K., Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects ofvowels in sets and order of training. In *Journal of Speech, Language, and Hearing Research*, 51(6), 1480-1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Odden, D. (2005). *Introducing Phonology*. Cambridge: Cambridge University Press.

Pruitt,S., Jenkins, J.J., Strange, W. (2006). Training the perception ofHindi dental and retroflex stops by native speakers of American English and Japanese. In *The Journal of the Acoustical Society of America*, 119(3), 1684-1696. https://doi.org/10.1121/1.2161427

Qian, M., Chukharev-Hudilainen, E., Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. In *Language Learning and Technology*, 22(1), 69-96. https://dx.doi.org/10125/44582

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rato, A., Rauber, A. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In the Scottish Consortium for ICPhS 2015. (Ed.), *Proceedings of the i8th International Congress of Phonetic Sciences*. Glasgow, UK: Glasgow University, 56.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., Munhall, K.G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. In *The Journal of the Acoustical Society of America*, 123(1), 397-413.https://doi.org/10.1121/1.2804942

Thomson, R.I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. In *CALICO Journal*, 28,744-765.

Thomson, R.I. (2012). Improving Li listeners' perception of English vowels: A computer-mediated approach. In *Language Learning*, 62, 1231-1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R.I. (2016a). Does training to perceive Li English vowels in one phonetic context transfer to other phonetic contexts? In *Canadian Acoustic*, 44(3), 198-199.

Thomson, R.I. (2016b). Learning to perceive English sounds using computer-assisted pronunciation training. *AL Forum*, November 2016, Alexandria, VA, TESOL International Association.

Thomson, R.L., Derwing, T.M. (2016). Is phonemic training using nonsense or real words more effective? In Levis, J., Le, H., Lucic, I., Simpson, E., Vo, S. (Eds.). *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*. Ames IA: Iowa State University, 88-97.

Wang, X., Munro, M.J. (2004). Computer-based training for learning English vowel contrasts. In*System*, 32(4), 539-552.

Wang,Y., Spence, M.M., Jongman, A., Sereno, J.A. (1999). Training American listeners to perceive Mandarin tones. In *The Journal of the Acoustical Society of America*, 106(6), 3649-3658. https://doi.org/10.1121/1.428217

Weber, A., Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. In *Journal of Memory and Language*, 50, 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0