

VOCE E APPRENDIMENTO MULTIMEDIALE

Anna De Meo, Marilisa Vitale, Elisa Pellegrino
Università degli Studi di Napoli "L'Orientale"
ademeo@unior.it, marilisavitale@hotmail.it, epellegrino@unior.it

1. SOMMARIO

Il presente studio ha indagato in maniera sperimentale l'impatto delle caratteristiche segmentali e soprasegmentali della voce sulla capacità degli apprendenti di elaborare e trattenere le informazioni in contesti di apprendimento a distanza. Per lo scopo una madrelingua italiana con pronuncia standard (priva di riconoscibili tratti regionali), due italiane native con accento regionale (campano e toscano), due parlanti non native di italiano con forte accento straniero (inglese e cinese) hanno letto un testo tecnico-scientifico sul tema della *speciazione* ovvero il processo di formazione di nuove specie zoologiche e botaniche, a partire da una di origine. Per includere nel corpus una voce priva di coloriture regionali a livello segmentale e/o soprasegmentale, ne è stata aggiunta una sesta sintetizzata che è stata anche manipolata attraverso il trapianto dell'intonazione e del ritmo delle cinque voci naturali. È stato così raccolto un corpus di parlato letto realizzato da 11 voci differenti: cinque voci naturali, una voce sintetizzata e cinque voci con segmenti sintetizzati e prosodia naturale.

In qualità di ascoltatori hanno partecipato alla sperimentazione 165 studenti campani, di età media pari a 21 anni e con un normale sviluppo delle funzioni uditive e cognitive. I partecipanti sono stati suddivisi in 11 gruppi omogenei per numerosità (15 ascoltatori per gruppo), provenienza, età, scolarizzazione, privi di conoscenze relative al tema della *speciazione*. Ciascun gruppo ha ascoltato la lezione sul tema della *speciazione* prodotta da una delle 11 voci selezionate per lo studio. Ai soggetti, che avevano dichiarato di non avere alcuna conoscenza dell'argomento, non è stato fornito alcun supporto visivo, né è stata data loro la possibilità di riascoltare il testo e/o di prendere appunti. Al termine dell'ascolto, è stato somministrato un test di comprensione orale, costruito con le tecniche della scelta binaria, dell'abbinamento parole/immagini e del riordino di sequenze di immagini. L'esercizio di scelta binaria è stato utilizzato solo come attività preparatoria di warm-up, finalizzata a predisporre positivamente i soggetti alle attività di verifica successive. Per questo motivo, gli esiti di questa prova non sono stati considerati nella valutazione del carico cognitivo.

Delle letture prodotte dalle cinque voci naturali e da quella sintetizzata è stata condotta un'analisi spettro-acustica. Per ciascuna produzione sono state misurate:

- il numero di sillabe realmente prodotte;
- in numero e la durata delle catene foniche prodotte (s);
- la durata delle pause silenziose e delle disfluenze (s);
- il valore minimo e massimo della frequenza fondamentale (f_0).

Sulla base di tali dati sono stati calcolati i seguenti indici ritmico-prosodici:

- velocità di articolazione (VdA) – sill/s;
- velocità di eloquio (VdE) – sill/s;
- fluenza (F) – sill/n. catene foniche;
- range tonale (RT) – st;
- composizione del parlato (% silenzi, tempo di fonazione, disfluenze).

Tali analisi hanno permesso sia di caratterizzare le sei produzioni orali in termini prosodico-intonativi, sia di accertare l'esistenza di una possibile correlazione tra il diverso profilo prosodico delle voci e gli esiti del test di comprensione.

Fatta eccezione dell'esercizio di warm-up, per il quale la percentuale di risposte corrette, come atteso, non ha mai superato la soglia della casualità, i risultati delle attività di abbinamento e di riordino hanno evidenziato un effetto positivo dell'utilizzo di una voce con pronuncia standard sull'elaborazione dei contenuti. Favorisce l'apprendimento anche l'esposizione all'accento regionale familiare all'ascoltatore, il campano. Non incontrando difficoltà particolari con la decodifica del segnale acustico, il ricevente riesce a canalizzare le sue risorse cognitive esclusivamente sulla strutturazione e la ritenzione delle informazioni. Particolarmente penalizzante ai fini dell'apprendimento è, invece, l'effetto prodotto dalla voce non nativa con un accento straniero poco familiare ai partecipanti; i gruppi a cui era stata somministrata la produzione della madrelingua cinese nella versione naturale e sintetizzata hanno infatti conseguito la percentuale superiore di risposte errate. Seppur in misura differenziata, deviazioni segmentali abbinata ad un'intonazione monotona e deviazioni prosodiche sovrapposte a segmenti sintetizzati sottraggono parte delle risorse cognitive necessarie all'attività di elaborazione dei contenuti, poiché l'apprendente deve impegnarsi nell'attività preliminare di decodifica del significante.

Le difficoltà incontrate dagli ascoltatori campani nella comprensione del parlato con accento cinese trovano riscontro nei dati dell'analisi spettro-acustica. La parlante cinese mostra valori di velocità di articolazione, di eloquio, fluidità e range tonale inferiori rispetto a quelli degli altri modelli analizzati (sintetizzato, non nativo e nativo).

Dal punto di vista della produzione, tali differenze denotano la maggiore difficoltà del locutore a gestire in maniera accurata il piano segmentale e soprasegmentale della L2. Dal punto di vista della ricezione, tali incertezze determinano nell'ascoltatore un sovraccarico cognitivo che ostacola la piena elaborazione e ritenzione delle informazioni.

2. INTRODUZIONE

Negli ultimi decenni le tecnologie dell'informazione e della comunicazione (TIC) hanno assunto un ruolo sempre più preponderante in ambito didattico (Chambers & Davies, 2001; Kervin & Derewianka, 2001). Con la diffusione delle tecnologie di rete e l'aumento significativo dell'e-learning (White, 2003), le TIC, inizialmente utilizzate solo come strumento di supporto all'insegnamento faccia a faccia, sono state adoperate in maniera sempre più consistente anche nei percorsi di apprendimento sviluppati in ambiente virtuale.

La diffusione delle procedure della formazione a distanza ha posto il problema di come integrare tali tecnologie nei percorsi di apprendimento di nuova generazione. Gli studi effettuati nell'ambito delle scienze cognitive hanno, infatti, dimostrato che i metodi seguiti nella progettazione della didattica online risultano talvolta inadeguati, poiché non calibrati sui processi mentali effettivamente coinvolti nell'elaborazione delle nozioni veicolate mediante supporti multimediali. Tali ricerche hanno sottolineato la necessità di sviluppare strumenti e percorsi di apprendimento che siano strutturati in funzione dei processi cognitivi implicati nell'elaborazione dei materiali didattici proposti e che tengano conto, quindi, dell'impatto delle tecnologie sui processi di apprendimento.

Una delle principali teorie elaborate a tale riguardo è quella del carico cognitivo (Chandler & Sweller, 1991; Jong de, 2010), secondo la quale l'utilizzo di materiale didattico rispondente alle esigenze del sistema cognitivo dello studente faciliterebbe

l'apprendimento. Se le risorse cognitive disponibili venissero indirizzate direttamente verso attività strettamente funzionali all'apprendimento, evitando il loro impiego, anche parziale, in elaborazioni preliminari e collaterali, esse verrebbero ottimizzate impedendo il sovraccarico cognitivo. Un'istruzione inefficace sarebbe, quindi, il risultato della necessità di integrare le informazioni necessarie all'esecuzione di un dato compito, con conseguente impossibilità di focalizzare l'attenzione su obiettivi specifici.

Tra gli elementi che possono interferire con l'elaborazione delle informazioni veicolate attraverso il canale multimediale è possibile includere la voce. Secondo uno studio di stampo psicopedagogico, condotto da Mayer & Moreno (2003), le voci utilizzate per lezioni multimediali possono alterare il processo e l'esito dell'apprendimento, in quanto lo sforzo cognitivo richiesto per decifrare un testo orale sottrarrebbe risorse mentali utili alla costruzione di collegamenti tra blocchi di informazione. L'utilizzo di una voce umana caratterizzata da accento standard, che non presenti alterazioni né sul piano segmentale né su quello soprasegmentale, sarebbe elaborata dal destinatario del messaggio più facilmente di una voce sintetizzata o con accento straniero/regionale e, di conseguenza, sarebbe più idonea a veicolare informazioni strutturate. I risultati dello studio, condotto sull'inglese americano, sembrano essere in linea con i principi della teoria del carico cognitivo. Tuttavia, va sottolineato che per la sperimentazione è stato considerato un unico accento straniero (voce inglese con accento russo) e un'unica voce sintetizzata.

3. LO STUDIO

Data la rilevanza del rapporto tra apprendimento e cognizione ai fini della progettazione della didattica in ambiente virtuale, il presente studio intende indagare in maniera sperimentale l'impatto delle caratteristiche segmentali e prosodiche di una voce impegnata nella lettura di un testo espositivo sulla capacità dell'ascoltatore di trattenere informazioni.

3.1. Il corpus

Il testo espositivo scelto per la sperimentazione è stato elaborato a partire da un articolo dell'enciclopedia online Treccani.it dedicato alla *speciazione*, ovvero al "processo di formazione di nuove specie zoologiche e botaniche, a partire da una di origine, attraverso l'insorgere di meccanismi di isolamento riproduttivo che impediscono lo scambio genetico tra le popolazioni" (<http://www.treccani.it/enciclopedia/speciazione/>). Il testo finale, frutto di un processo di semplificazione e adeguamento alla fruizione con la sola modalità uditiva, è costituito da 925 parole. La rielaborazione testuale è stata resa necessaria dalla presenza di un numero elevato di tecnicismi e dalla contestuale assenza di momenti di sintesi, che avrebbero potuto alterare l'esito dello studio: i contenuti del testo avrebbero richiesto, infatti, un'elaborazione dell'informazione eccessivamente dispendiosa per soggetti non esperti, rendendo così difficile la valutazione del carico cognitivo operato da elementi non testuali. La scelta del tipo di testo è stata motivata dalla necessità di simulare una reale condizione di studio, con un tema poco noto agli studenti di lingue, ma presente nei programmi di scienze delle scuole medie di secondo livello. Il testo è stato giudicato fruibile anche per una didattica scolastica da tre insegnanti di scienze di istituti superiori campani.

3.2. Le voci

Le voci sono state selezionate sulla base di un test preliminare, che ha coinvolto 13 locutrici, tra cui:

- 2 italiane native con pronuncia standard, ossia priva di riconoscibili tratti regionali;

- 7 italiane native con accento regionale (3 campane, 1 piemontese, 1 emiliana, 1 toscana, 1 lucana);
- 4 parlanti non native di italiano (2 inglesi e 2 cinesi).

Il test percettivo, somministrato a 30 ascoltatori campani (maschi e femmine, età media 30 anni), è servito a selezionare una voce con accento standard, due voci con accento regionale e due con accento straniero. Per quanto riguarda la selezione delle voci con accento regionale e con accento straniero, la scelta è stata operata per rapporto a due parametri:

a) il grado di accento percepito, preferendo le voci valutate come portatrici di un accento molto marcato e, di conseguenza, più idonee a verificare un eventuale effetto del carico cognitivo a questo associato;

b) il grado di familiarità degli ascoltatori con gli accenti considerati, allo scopo di evidenziare possibili differenze legate al maggiore o minore impegno delle risorse cognitive nella decodifica del segnale acustico.

Nella fase di pre-test, agli ascoltatori è stato somministrato un estratto di 80 parole tratto dal testo della *speciazione*, prodotto dalle 13 voci femminili individuate. Per ciascuna di esse, è stato chiesto di indicare:

- la tipologia di accento (nativo/non nativo);
- il grado di accento regionale/straniero su una scala a cinque punti;
- solo per le voci native, il grado di vicinanza rispetto alla propria varietà di italiano su una scala a cinque punti.

In base ai risultati del pre-test, i soggetti selezionati per la sperimentazione sono stati:

- 1 italiana nativa con pronuncia standard (Sta)
- 2 italiane native con accento regionale forte
 - una con accento campano (il più vicino all'accento degli ascoltatori) (Cam)
 - una con accento toscano (il più lontano dall'accento degli ascoltatori) (Tos);
- 2 parlanti non native di italiano, entrambe con forte accento straniero
 - una con accento inglese definito familiare (Ing);
 - una con accento cinese definito non familiare (Cin).

Alle cinque voci femminili naturali, selezionate attraverso il pre-test percettivo, ne è stata aggiunta una sesta sintetizzata, ottenuta mediante il sistema di conversione text-to-speech supportato dal sistema operativo Mac OS X Mavericks (10.9.2), per avere nel campione da testare anche una voce femminile che non presentasse tratti segmentali e/o soprasegmentali riconducibili a una particolare varietà di italiano.

3.2.2. Manipolazione delle voci

Come ampiamente riportato in letteratura, i tratti caratterizzanti i diversi accenti sono di duplice natura, ossia legati a una diversa articolazione dei segmenti o a una diversa struttura melodica dell'enunciato. I testi registrati dai cinque soggetti selezionati, pertanto, sono stati manipolati al fine di valutare in maniera puntuale il diverso impatto degli elementi segmentali e prosodici sul carico di lavoro mentale necessario per l'elaborazione del testo espositivo considerato.

Mediante la tecnica del trapianto ritmico-prosodico (Pettorino & Vitale, 2012; Yoon, 2007), che sfrutta l'algoritmo PSOLA (Moulines & Charpentier, 1990) implementato nel software Praat (Boersma & Weenink, 2012), l'intonazione e il ritmo delle cinque voci natu-

rali sono stati sovrainposti ai segmenti prodotti dalla voce artificiale. Sono state quindi ottenute le seguenti voci manipolate:

- voce 1: segmenti sintetizzati + prosodia dell'italiano nativo con pronuncia standard (StaP);
- voce 2: segmenti sintetizzati + prosodia dell'italiano con accento campano (CamP);
- voce 3: segmenti sintetizzati + prosodia dell'italiano con accento toscano (TosP);
- voce 4: segmenti sintetizzati + prosodia dell'italiano con accento inglese (IngP);
- voce 5: segmenti sintetizzati + prosodia dell'italiano con accento cinese (CinP).

3.3. Il test

L'esistenza di una correlazione tra voce impiegata nell'esposizione di un testo e apprendimento significativo del contenuto veicolato è stata verificata attraverso un test di comprensione orale.

Sono stati selezionati 165 italiani nativi, suddivisi in 11 gruppi, omogenei per numero, provenienza, età, scolarizzazione e competenza linguistica. A ciascun gruppo, composto da 15 soggetti, è stato affidato il compito di ascoltare una breve lezione sul tema della *speciazione* e di svolgere, al termine dell'ascolto, tre brevi esercizi di verifica delle conoscenze acquisite. I partecipanti, maschi e femmine, età media 21 anni, tutti studenti universitari di lingue e letterature straniere presso l'Università di Napoli L'Orientale, nati e residenti in Campania, privi di deficit uditivi, hanno dichiarato di non possedere alcuna conoscenza preliminare relativa al tema della *speciazione*, di avere una competenza intermedia della lingua inglese ma di non aver mai studiato il cinese. La partecipazione all'esperimento è stata esclusivamente su base volontaria. La composizione dei vari gruppi di ascoltatori, omogenei sulla base di numerosi fattori sociolinguistici, è stata pensata con l'obiettivo di minimizzare la variabilità intergruppo.

L'attività di ascolto è avvenuta in maniera individuale e in cuffia, per evitare distrazioni o problemi legati all'amplificazione più o meno adeguata dell'audio. Inoltre, ai soggetti non è stato fornito alcun supporto visivo e non è stata data la possibilità né di riascoltare il testo né di prendere appunti. A ogni gruppo è stata assegnata e somministrata solo una delle 11 voci selezionate per lo studio (5 naturali, 1 sintetizzata, 5 manipolate), poiché la somministrazione ripetuta dello stesso testo, sebbene prodotto da voci diverse, avrebbe provocato un effetto di familiarizzazione e di accumulo dell'informazione, che avrebbe alterato gli esiti dello studio.

Il test di verifica somministrato al termine dell'ascolto ha previsto:

- 1 questionario a scelta binaria (5 domande di difficoltà crescente)
- 1 attività di abbinamento parole/immagini (4 item)
- 1 attività di riordino di sequenze di immagini (4 item).

Gli esiti del questionario a scelta binaria non sono stati considerati nella valutazione del carico cognitivo, poiché questa attività ha una probabilità di risposte casuali pari al 50%, ed è stata utilizzata esclusivamente come warm-up, allo scopo di predisporre positivamente i soggetti alle attività di verifica successive e abbassare il filtro affettivo. L'insorgere dell'ansia da prestazione avrebbe probabilmente comportato un numero di risposte non corrette superiore a quello atteso, condizionando così l'esito dell'intero test. La difficoltà crescente delle cinque domande previste per questa prima fase ha gradualmente introdotto i soggetti alla parte centrale della verifica, costituita dalle attività di riordino e abbinamento. Queste erano basate su immagini decontestualizzate, non incontrate prima dai soggetti, e sottintendevano la transcodifica dei contenuti da un testo lineare di natura fonico-uditiva a

uno non lineare di natura grafico-visiva. Questi due esercizi, particolarmente impegnativi sul piano cognitivo (Cummins, 1984), hanno permesso di valutare in maniera più oggettiva quanto dell'input orale si fosse trasformato in reale intake.

3.4. Analisi acustiche

Le cinque voci naturali e la voce sintetizzata sono state sottoposte ad analisi spettroacustica, al fine di verificare se eventuali differenze ritmico-prosodiche potessero corrispondere a dissimilarità nell'esecuzione della prova di comprensione orale.

Per ciascun testo orale prodotto dalle 6 voci sono state rilevate le seguenti misure:

- il numero di sillabe realmente prodotte;
- in numero e la durata in secondi delle catene foniche prodotte, ossia delle porzioni di parlato comprese tra due silenzi;
- la durata in secondi delle pause silenziose e delle disfluenze;
- il valore minimo e massimo della frequenza fondamentale (f_0).

Sulla base di tali dati sono stati calcolati i seguenti indici ritmico-prosodici:

- velocità di articolazione (VdA) – sill/s;
- velocità di eloquio (VdE) – sill/s;
- fluenza (F) – sill/n. catene foniche;
- range tonale (RT) – st;
- composizione del parlato (% silenzi, tempo di fonazione, disfluenze).

4. RISULTATI

4.1. Il test

Come atteso, il questionario a scelta binaria ha ottenuto una media di risposte corrette pari al 50% circa per ciascuna delle voci utilizzate, mantenendosi quindi in un intervallo di valori attribuibili alla casualità (figura 1).

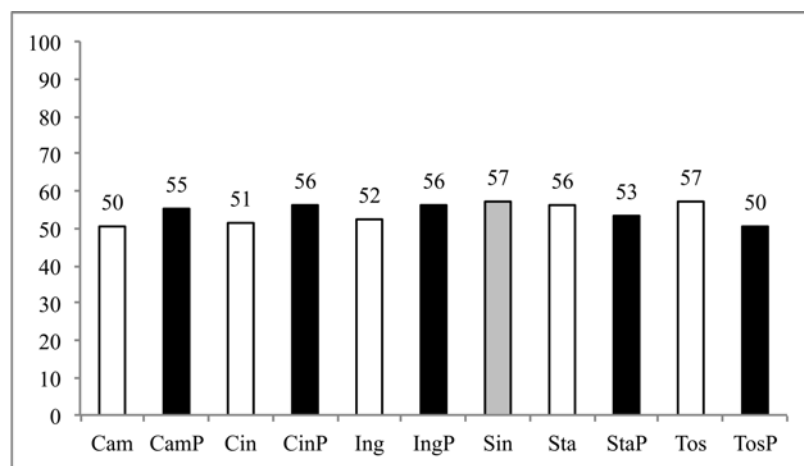


Figura 1: Questionario a scelta binaria - percentuale di risposte corrette.

Per le attività di abbinamento e riordino di immagini i dati sono stati analizzati scorpondo i soggetti che hanno effettuato correttamente l'intero esercizio (4/4 = 4 item su 4 ab-

binati o riordinati correttamente) e quelli che lo hanno sbagliato completamente (0/4 = 0 item su 4 abbinati o riordinati correttamente).

La prima delle due attività principali somministrate, l'abbinamento di immagini e parole (figura 2), evidenzia un effetto positivo dell'utilizzo di una voce con pronuncia standard sull'elaborazione dei contenuti di un testo orale, confermando quanto rilevato dallo studio di Mayer e Moreno (2003). Un aspetto interessante messo in evidenza dai risultati del test è che la voce con accento familiare agli ascoltatori (campano) sembra non interferire negativamente con il processo di decodifica e comprensione del messaggio. L'ascoltatore, elaborando rapidamente il segnale acustico familiare, concentra le risorse mentali a propria disposizione sulla strutturazione e l'accumulo delle informazioni.

Particolarmente negativo, per contro, è l'effetto prodotto da una voce con accento straniero scarsamente noto ai partecipanti (cinese). La voce naturale con accento cinese, infatti, corrisponde al numero più elevato di test totalmente errati. La combinazione di segmenti poco accurati o erroneamente articolati e di una prosodia non adeguata e prevalentemente monotona ha richiesto agli ascoltatori un maggiore impegno nella decodifica del messaggio, sottraendo buona parte delle risorse mentali all'attività di elaborazione dei contenuti. Ciò sembra essere confermato dai risultati ottenuti dalla voce manipolata, ottenuta con prosodia cinese e segmenti sintetizzati: riducendo il numero di tratti vocalici e/o consonantici errati, è stata favorita una più rapida decodifica del testo, che ha consentito l'immagazzinamento di un maggior numero di informazioni.

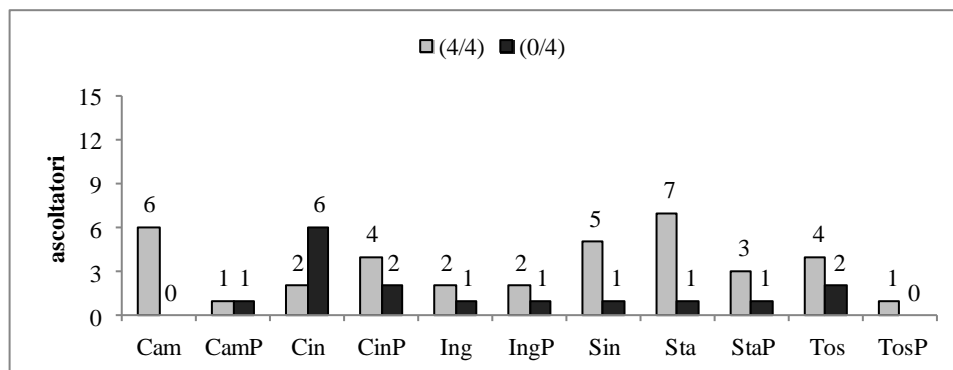


Figura 2: Abbinamento parole/immagini – numero di risposte corrette.

A differenza di quanto avviene per la manipolazione della voce cinese, in tutti gli altri casi, l'utilizzo di una prosodia naturale sovrapposta a segmenti sintetizzati produce un effetto negativo sull'ascoltatore, come si evince dalla figura 2. Ciò vale soprattutto nel caso delle voci native (standard, campana e toscana), per le quali la manipolazione determina un'alterazione del segnale acustico tale da interferire con l'elaborazione del messaggio. Non si osservano variazioni nel caso della voce inglese.

I risultati relativi all'esercizio di riordino di sequenze di immagini (figura 3) confermano la tendenza osservata per l'attività precedente. La voce standard e quella con accento campano, familiare agli ascoltatori, anche in questo caso, favoriscono una più agevole decodifica del testo e permettono di ottenere il maggior numero di test completati in maniera corretta. Tuttavia, per la seconda attività si osserva un'incidenza inferiore della manipola-

zione del segnale fonico, probabilmente riconducibile all'esecuzione del task immediatamente precedente, che ha creato una sorta di agevolazione dell'elaborazione cognitiva.

Tale processo giustificerebbe anche l'assenza di miglioramenti relativi alla manipolazione della voce cinese, per la quale si notano risultati negativi sia nella versione naturale sia in quella sintetizzata.

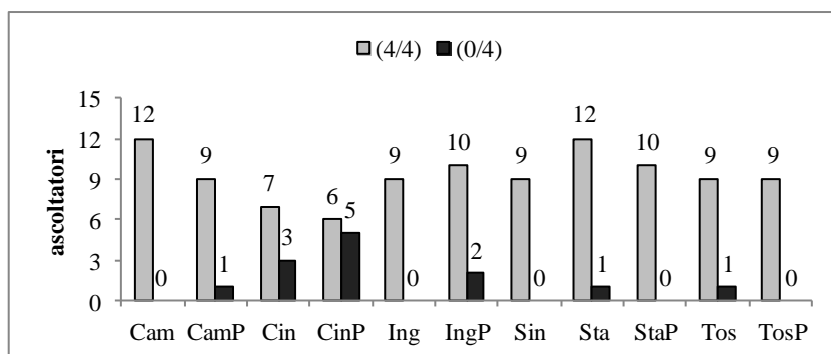


Figura 3: Riordino di sequenze di immagini – numero di risposte corrette.

4.2. Analisi acustiche

Le difficoltà incontrate dagli ascoltatori campani nella decodifica del parlato con accento cinese trova riscontro nei dati dell'analisi spettro-acustica (tabella 1).

	Fluenza (sill/n. cf)	VdA (sill/s)	VdE (sill/s)	Range tonale (st)	% silenzi	% tempo fonazione	% disfluenze
Sin	28.3	6.8	6.6	10.6	96	4	0
Sta	17.0	5.7	5.2	12.7	91	9	0
Cam	16.8	5.6	5.0	11.7	90	10	0
Tos	14.0	5.9	5.2	9.2	89	11	0
Cin	7.8	4.5	3.6	6.3	80	20	0
Ing	14.3	5.1	4.6	9.2	89	11	0

Tabella 1: Risultati dell'analisi spettro-acustica – valori medi.

La parlante cinese mostra valori marcatamente differenti rispetto a quelli degli altri soggetti analizzati (voce sintetizzata, voci non nativa inglese e voci native) per tutti gli indici prosodici considerati. Una più bassa velocità di articolazione può essere indicativa di una maggiore difficoltà incontrata dalla locutrice nella realizzazione del gesto articolatorio. Una particolare incertezza nella gestione del testo, sebbene si tratti di parlato letto, si evince invece dalla contenuta velocità di eloquio e dal moderato valore della fluenza. Questi due indici, infatti, sono strettamente collegati al numero e alla durata delle pause silenziose e rivelano la necessità di interrompersi più frequentemente. La percentuale di silenzi prodotta dalla cinese è, in effetti, più alta rispetto a quella riscontrata nelle produzioni degli altri soggetti.

Per quanto riguarda le variazioni intonative, il range tonale più ristretto, soprattutto se confrontato con quello prodotto dai nativi, rispecchia una problematicità nella modulazione dell'andamento intonativo: picchi tonali meno marcati corrispondono verosimilmente a una

mancanza di enfasi nella lettura che, se presente, aiuta l'ascoltatore a focalizzare l'attenzione sui punti chiave del testo, agevolandone la comprensione.

CONCLUSIONI

Lo studio condotto ha permesso di confermare la validità della teoria del carico cognitivo, attraverso un esperimento che ha utilizzato il solo canale vocale per la trasmissione di un testo accademico scientifico, che è stato elaborato uditivamente senza alcun supporto visivo. Il testo, costituito da 925 parole, ha proposto agli ascoltatori, 165 studenti di lingue, il tema della *speciazione*, del quale essi hanno dichiarato di non avere alcuna conoscenza pregressa.

Le voci utilizzate per lo studio avevano diversi gradi di accento, regionale e straniero, e mostravano diversi gradi di vicinanza o familiarità alla varietà di italiano degli ascoltatori, tutti campani. Alle voci naturali (1 standard, 1 campana, 1 toscana, 1 cinese, 1 inglese), tutte femminili, sono state aggiunte una voce sintetizzata e cinque voci manipolate, ottenute sovrapponendo alla voce sintetizzata la prosodia delle cinque voci naturali.

Per la verifica della comprensione orale sono stati utilizzati tre task (scelta binaria, abbinamento parole/immagini, riordino di immagini) che hanno mostrato gradi di difficoltà variabili e una diversa influenza delle voci sul carico cognitivo.

L'attività di abbinamento parole/immagini ha evidenziato un effetto positivo sia della voce standard sia di quella con accento regionale familiare (campano), mentre la voce con accento straniero forte e minore grado di familiarità (cinese) ha inciso negativamente sui risultati. Situazioni intermedie sono attestate per la voce regionale poco familiare (toscana) e quella straniera più familiare (inglese). La manipolazione delle voci ha prodotto effetti generalmente negativi, tranne nel caso della voce sintetizzata con prosodia cinese, che ha interferito in maniera minore di quella naturale sull'elaborazione dei contenuti veicolati.

Nell'attività di riordino delle sequenze di immagini si confermano le tendenze già osservate, anche se in maniera meno marcata. Per quanto riguarda la manipolazione delle voci, in questo caso non ha inciso in maniera rilevante sui risultati, che sono piuttosto stabili, anche per la voce cinese. Una probabile motivazione di tale fenomeno potrebbe essere la familiarizzazione con le immagini avvenuta nel task immediatamente precedente.

I risultati del presente studio suggeriscono la necessità di una particolare attenzione nella selezione delle voci da utilizzare nella costruzione di materiale didattico multimediale o percorsi di formazione a distanza. Voci con accento standard o con accento regionale familiare per gli ascoltatori sembrerebbero creare le condizioni ottimali per l'elaborazione di contenuti trasmessi attraverso il canale uditivo, poiché esse non richiedono l'impiego di risorse cognitive specifiche che verrebbero sottratte al processo di strutturazione delle informazioni. Ovviamente, nel caso della voce con accento regionale il prodotto costruito per la didattica a distanza non potrebbe avere la stessa diffusione territoriale di quello con accento standard. Da evitare, invece, voci con accento straniero forte, o con accento regionale percepito come distante dalla varietà linguistica degli ascoltatori.

BIBLIOGRAFIA

Boersma, B. & Weenink, D. (2012), Praat: doing phonetics by computer [Computer program]. Version 5.3.15 retrieved 22 May 2012 from <http://www.praat.org/>.

- Chambers, A. & Davies, G. (2001), *Information and Communications Technologies in Language Learning: A European Perspective*, Lisse: Swets & Zeitlinger.
- Chandler, P. & Sweller, J. (1991), Cognitive Load Theory and the Format of Instruction, *Cognition and Instruction*, 8(4), 293-332.
- Cummins, J. (1984), *Bilingualism and special education: issues in Assessment and Pedagogy*, Clevedon UK: Multilingual Matters.
- Jong de, T. (2010), Cognitive load theory, educational research, and instructional design: some food for thought, *Instructional Science*, 38, 105-134.
- Kervin, L. & Derewianka, B. (2011), New technologies to support language learning, in *Materials Development in Language Teaching* (B. Tomlinson, editor), Cambridge, UK: Cambridge University Press.
- Mayer, R.E. & Moreno, R. (2003), Nine Ways to Reduce Cognitive Load in Multimedia Learning, *Educational Psychologist*, 38(1), 43-52.
- Moulines, E. & Charpentier, F. (1990), Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones, *Speech Communication*, 9, 453-467.
- Pettorino, M. & Vitale, M. (2012), Transplanting native prosody into second language speech, in *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, (M.G. Busà & A. Stella, editors), Padova: CLEUP, 95-99.
- Yoon, K. (2007), Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody, *Journal of the Modern British & American Language & Literature*, 25(4), 197-215.
- White, C. (2003), *Language learning in distance education*, Cambridge: Cambridge University Press.