

Possibilità di utilizzo delle tecnologie vocali nella riabilitazione dei disturbi semantico-lessicali nell'afasia¹

Morena Danieli²

Il mio intervento esplora la possibilità di utilizzare nella riabilitazione dell'afasia le tecnologie di sintesi, riconoscimento e comprensione del linguaggio naturale. Si tratta di tecnologie sviluppate negli ultimi trent'anni nell'ambito dell'intelligenza artificiale con l'obiettivo di dotare i calcolatori di competenze linguistiche volte a facilitare l'interazione tra gli esseri umani e le macchine. Il perfezionamento dei modelli e degli algoritmi e la crescente potenza di calcolo dei dispositivi elettronici hanno reso le tecnologie vocali sempre più flessibili, precise, adatte ad essere utilizzate nella vita quotidiana in applicazioni utilizzate nella telefonia mobile, nella realtà virtuale, nell'erogazione automatica di informazioni e, nel caso del riconoscimento del parlante, anche in applicazioni di sicurezza e di *intelligence*. Tuttavia nel campo della riabilitazione, in particolare nella riabilitazione dell'afasia, i progetti di integrazione di tali tecnologie sono sporadici, nonostante l'afasia sia in Italia e all'estero un problema sanitario di particolare rilievo. Le stime di incidenza e prevalenza pubblicate in anni recenti riferiscono percentuali di invalidità per afasia oscillanti tra il 21 e il 38% nei soggetti colpiti da ictus³.

Chi si occupa di afasia sa quanto sia urgente poter disporre di interventi specifici e di strumenti flessibili, personalizzabili e di provata efficacia per la riabilitazione. Gli operatori della riabilitazione sono altrettanto consapevoli di come quest'esigenza sia in conflitto con la sproporzione drammatica attuale tra risorse a disposizione e domanda crescente di cura e riabilitazione. Una delle variabili cruciali del processo di riabilitazione dell'afasia è l'intensità dell'intervento riabilitativo, ma questo aspetto può essere sottoposto a severe limitazioni dovute alla carenza di risorse. L'integrazione tra le sedute di riabilitazione tradizionali e interventi assistiti da sistemi capaci di comprendere il parlato del soggetto con afasia potrebbe contribuire, sia pure per alcuni specifici compiti riabilitativi, ad aumentare l'intensità della cura. Inoltre è possibile ipotizzare che l'interazione personalizzata con sistemi interattivi dedicati alla riabilitazione possa

¹ Testo presentato al Simposio Nazionale *Riabilitare la comunicazione. Tecniche e Tecnologie Applicate*, Fondazione Molo e Università degli Studi di Torino, 29-30 novembre 2013.

² Morena Danieli è psicologa con formazione in neuroscienze cognitive, psicoterapeuta con orientamento psicoanalitico, *research fellow* presso il *Signals and Interactive Systems Lab* del Dipartimento di Ingegneria e Scienza dell'Informazione dell'Università degli Studi di Trento (morena.danieli@unitn.it).

³ Si vedano gli studi epidemiologici del "Progetto afasia post-ictus" sponsorizzato da A.I.T.A. e coordinato dal Dipartimento di Farmacologia Clinica e Epidemiologica del Consorzio Mario Negri Sud (http://www.aitafederazione.it/ricerca_epidemiologica.php).

contribuire ad accrescere nel paziente afasico la disponibilità a investire tempo, perseveranza e coerenza nel processo di recupero.

Nel valutare la possibilità di utilizzare le tecnologie di sintesi e riconoscimento della voce nella riabilitazione dell'afasia è necessario rispondere ad alcune domande importanti. Innanzitutto, se il parlato afasico possa essere compreso da un sistema automatico di riconoscimento in modo sufficientemente affidabile. In secondo luogo sarà necessario chiedersi quali siano le competenze linguistiche e comunicative che possono giovare maggiormente di un programma riabilitativo che utilizzi il riconoscimento del parlato. In terzo luogo bisognerà valutare quali siano i vantaggi per il paziente e quali quelli per il terapeuta. Nelle pagine seguenti proporrò alcune risposte iniziali a queste domande che considero preliminari per poter avviare progetti di integrazione tra tecnologie del parlato e tecniche di riabilitazione dell'afasia.

Applicazioni di *speech processing* nel campo della riabilitazione

Nella letteratura scientifica sul riconoscimento automatico del linguaggio il termine *speech processing* indica lo studio dei segnali vocali e le tecniche di analisi di tali segnali. Lo *speech processing* include numerose aree di indagine, tra le quali spiccano il riconoscimento, chiamato talvolta anche “riconoscimento della voce”, e la sintesi del parlato. Il riconoscimento del parlato ha l'obiettivo di decifrare il contenuto linguistico del segnale vocale e di trasformarlo in un formato rappresentabile nei computer. La sintesi della voce è la produzione artificiale di linguaggio parlato. Semplificando un poco possiamo dire che mentre un riconoscitore del parlato traduce il segnale vocale in parole e simboli (comprensione), un sistema di sintesi della voce produce dei segnali vocali a partire da un testo (produzione).

Le prime applicazioni di *speech processing* nel campo della riabilitazione risalgono alla seconda metà degli anni '80 del novecento. Si trattava di applicazioni pionieristiche di sintesi della voce rivolte a soggetti non vedenti o ipovedenti. I lettori automatici di testi potevano offrire a queste persone una maggiore accessibilità all'informazione scritta. Nei decenni seguenti la sintesi della voce è stata utilizzata anche in applicazioni volte ad accrescere la possibilità di comunicare di soggetti con gravi disturbi della fonazione, come nel caso dei *voice output communication aids* utilizzati dai soggetti con SLA.

Oggi numerose applicazioni di realtà virtuale, che talvolta includono agenti conversazionali *embodied*, sono utilizzate in programmi di riabilitazione. Alcuni sistemi utilizzano la sintesi della voce per generare i messaggi comunicati dall'agente conversazionale, ma quasi nessuno sfrutta il riconoscimento del parlato per decodificare e comprendere i messaggi comunicati dal paziente. Credo che ciò sia in gran parte dovuto alla maggiore difficoltà del processo di analisi e comprensione automatica del segnale vocale, e alle conseguenti prestazioni imperfette dei sistemi di riconoscimento della voce, a fronte della precisione, naturalezza timbrica e immediatezza d'uso della sintesi della voce.

Il riconoscimento del parlato

Per riconoscere e comprendere automaticamente il linguaggio sono necessarie numerose fasi di analisi, illustrate schematicamente nella Figura 1.

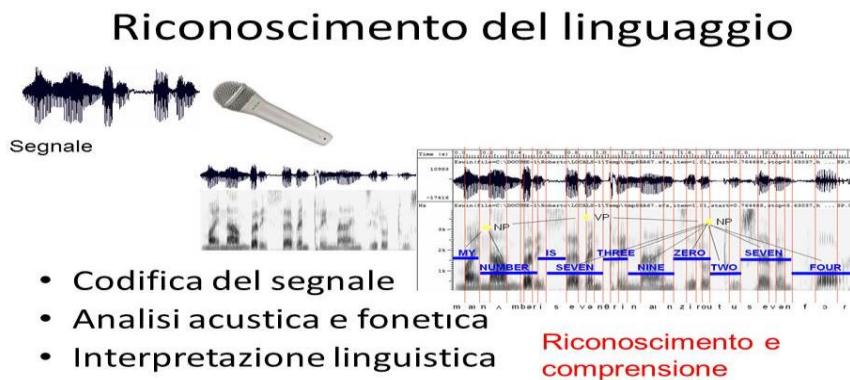


Figura 1

A sinistra nella figura possiamo vedere che il segnale acustico deve essere campionato e codificato, cioè convertito da segnale analogico a digitale, per poter essere rappresentato e trattato nel calcolatore. Il segnale viene quindi analizzato innanzitutto per separare il suono linguistico dal silenzio e/o dal rumore di fondo e ambientale. Il passo successivo è costituito dall'analisi acustica e fonetica del segnale con l'obiettivo di riconoscere, nella variabilità della produzione orale, i tratti distintivi dei fonemi, le parole e i loro confini. L'interpretazione linguistica ricostruisce la struttura morfologica delle parole pronunciate, la struttura sintattica della frase, la struttura predicativa e l'interpretazione semantica. Infine, l'analisi pragmatica filtra i fenomeni tipici dell'esecuzione, per esempio le disfluenze, e interpreta i fenomeni conversazionali. Il risultato del processo di analisi varia a seconda delle applicazioni: in alcuni casi il sistema restituisce una stringa di parole, in altri casi una struttura semantica che rappresenta il significato della frase pronunciata.

Ciascuna fase di analisi è caratterizzata da difficoltà di vario genere che sorgono quando gli algoritmi devono far fronte alla variabilità della lingua, una delle caratteristiche più affascinanti del linguaggio umano, che si manifesta tra i parlanti e nel singolo parlante.

La variabilità tra i parlanti è dovuta alle caratteristiche individuali del *pitch* e della curva intonativa, oltre che all'accento regionale, alla realizzazione allofonica e alle qualità timbriche specifiche delle diverse voci. Ma anche il parlato di un singolo soggetto può variare nel tempo a seconda dello stato di salute, delle emozioni e della situazione ambientale. Esiste infine una variabilità legata allo stile di parlato: spontaneo o letto, familiare o formale.

La variabilità costituisce la sfida principale per le tecnologie di riconoscimento della voce, ma poiché tali sistemi sono in grado di “apprendere”, sono talvolta predisposti per potersi adattare al singolo parlante. Infatti, i sistemi di comprensione del parlato più recenti si basano sull’apprendimento automatico e sono capaci di svolgere il loro compito integrando informazioni di natura diversa: non solo le caratteristiche fonetiche, ma anche quelle soprasegmentali, i segnali fisiologici e un insieme limitato di dati comportamentali. Il loro nucleo centrale è costituito da un sistema di apprendimento automatico a partire da esempi.

L’apprendimento automatico è alla base dei cosiddetti riconoscitori acustico-fonetici “dipendenti dal parlante”, basati su algoritmi di *training* che consentono di aumentare la loro tolleranza verso le idiosincrasie individuali della produzione linguistica. Il parlante deve pronunciare alcune parole e frasi predefinite affinché il sistema acquisisca esempi caratteristici della voce individuale. Circa trenta minuti di parlato individuale sono sufficienti per allenare il sistema. Si tratta di sistemi più robusti rispetto a quelli indipendenti dal parlante, ma sono difficilmente utilizzabili nella riabilitazione dell’afasia perché tendono ad adattarsi agli errori del paziente. Al contrario, i sistemi indipendenti dal parlante sembrerebbero più adatti per le applicazioni di riabilitazione: abbastanza flessibili nel tollerare la variabilità fonetica, dovrebbero essere particolarmente curati per quanto riguarda la copertura linguistica in ambito conversazionale.

***Speech processing* e riabilitazione dell’afasia**

Vediamo ora quali aspetti della riabilitazione possono trarre giovamento dall’utilizzo di tecnologie di riconoscimento e comprensione della voce.

Nelle linee guida della Federazione dei Logopedisti Italiani⁴ ho trovato almeno tre raccomandazioni che possono riguardare il nostro tema. La prima raccomandazione riguarda gli interventi di riabilitazione basati sul computer: questi interventi, si dice, sono consigliati a supporto e non in sostituzione del trattamento logopedico⁵. La seconda raccomandazione riguarda il monitoraggio dei risultati della riabilitazione al fine di calibrarne la durata, sottolineando come sia importante correlare la durata del trattamento alla verifica di efficacia mediante la valutazione e il monitoraggio costante dei risultati⁶. La terza raccomandazione stabilisce che “è indicato un trattamento intensivo”⁷.

L’ipotesi che propongo è conforme a tali raccomandazioni perché prevede che il sistema automatico funzioni a supporto del terapeuta (Raccomandazione 25), consenta di

⁴ Federazione Logopedisti Italiani (a cura di), *Linee guida sulla gestione logopedica del paziente afasico adulto*, 2009; disponibile in http://www.snlg-iss.it/linea_guida_fli_afasia

⁵ Raccomandazione 25, p. 63, *ibidem*.

⁶ Raccomandazione 2.41, p. 76, *ibidem*.

⁷ Raccomandazione 2.43, p. 77, *ibidem*.

monitorare i risultati degli esercizi (Raccomandazione 2.41), fornendo un supporto efficace per aumentare l'intensità dell'intervento riabilitativo (Raccomandazione 2.43). Credo inoltre che l'utilizzo di sistemi basati su tecniche di intelligenza artificiale possa favorire l'assunzione di un ruolo attivo da parte del paziente nel gestire a casa una parte del programma riabilitativo, scegliendo i tempi più opportuni della giornata in cui dedicarsi agli esercizi, potendo tuttavia disporre della valutazione del terapeuta anche sul lavoro svolto autonomamente.

Quali compiti riabilitativi possono essere svolti con l'assistenza dei sistemi di riconoscimento e comprensione della voce? Per sceglierli dobbiamo fare riferimento a un modello neuropsicologico. Il modello a cui mi riferisco è il cosiddetto "modello a due vie"⁸, un modello largamente, anche se non universalmente, condiviso. In base a questo modello si ipotizza che l'elaborazione delle parole avvenga in virtù di informazioni relative al significato separate dalle informazioni relative alla pronuncia o alla scrittura delle parole. Mentre le informazioni relative al significato sono rappresentate in una componente indipendente dalla forme di parola fonetiche o ortografiche, le informazioni relative rispettivamente all'ortografia e alla pronuncia sono rappresentate nei lessici ortografici e fonologici di input e di output. I primi supportano il riconoscimento delle parole, i secondi la loro produzione. Il danno a una o più componenti lessicali determina una gamma di deficit funzionali nella produzione o nel riconoscimento delle parole. La maggior parte degli interventi riabilitativi dell'afasia prevede, laddove appropriato, una rieducazione delle componenti lessicali e del sistema semantico.

Lo stato attuale della tecnologia di riconoscimento della voce consente di poter svolgere con l'ausilio di tale tecnologia i compiti volti alla riabilitazione dei deficit semantico-lessicali, in particolare dei disturbi determinati dal danno a una componente lessicale coinvolta nell'elaborazione delle parole. Più precisamente, ritengo che sia possibile sfruttare i sistemi di riconoscimento vocale per riabilitare il sistema semantico e i lessici di output, il cui danno si manifesta spesso con disturbi di rievocazione delle parole.

I lessici fonologici di output sono attivati quando denominiamo un oggetto, quando parliamo spontaneamente, quando leggiamo o ripetiamo singole parole. Per la loro riabilitazione sono state avanzate diverse proposte terapeutiche, che vanno dai classici esercizi di denominazione, all'evocazione lessicale per fonema o per categoria semantica, fino agli esercizi con il supporto del vocabolario.

⁸ Cfr. A. Basso, *Aphasia and its Therapy*, Oxford, Oxford Univ. Press, 2003; A. Caramazza, "Some aspects of language processing revealed through the analysis of acquired aphasia: the lexical system", *Ann Rev Neurosci* 1988, vol. 11, pp. 395-421; P. Marangolo, A. Basso, "La riabilitazione dei deficit lessicali", in A. Mazzucchi (a cura di), *La riabilitazione neuropsicologica*, Masson, Milano, 2006, pp. 55-70; Patterson, K. et al. "Facilitation of word retrieval in aphasia", in Code, C. & Muller, D. (a cura di), *Aphasia Therapy*, E. Arnold, London, 1983, pp. 264-366;

Un esempio: la riabilitazione dei lessici fonologici nel progetto Vithea

Vithea⁹ è un progetto nato dalla collaborazione tra il centro studi portoghese INESC e la Facoltà di Medicina dell'Università di Lisbona. L'obiettivo del progetto è la realizzazione di una piattaforma online che consenta al terapeuta di creare esercizi per il singolo paziente, o per gruppi di pazienti, e di monitorarne i risultati. La piattaforma consente inoltre al paziente di collegarsi, via web, al sistema e di eseguire gli esercizi con il supporto di un agente conversazionale che fornisce un feedback relativamente ai risultati delle singole sedute di riabilitazione. L'interfaccia per il modulo paziente, che vediamo schematizzato nella Figura 2, è disponibile anche per l'utilizzo con un dispositivo mobile.

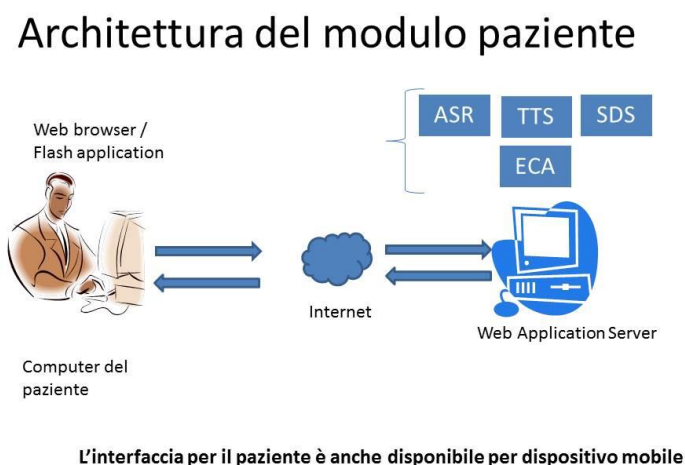


Figura 2

Il focus terapeutico del progetto Vithea è il recupero delle funzioni linguistiche correlate alla denominazione. Gli esercizi contengono stimoli semantici presentati in diverse modalità e contesti. Il protocollo prevede esercizi di training che presentano in modo casuale e attraverso varie modalità (grafica, uditiva e pittorica) stimoli che il paziente è invitato a denominare ad alta voce. L'output prodotto dal paziente è registrato e valutato dal sistema. Il paziente riceve quindi un feedback, può riascoltare la propria risposta e, in caso di errore, può ripetere l'esercizio.

Il sistema di riconoscimento vocale (ASR nello schema in Figura 2) ha il compito di decodificare, trascrivere e valutare la risposta del paziente. La verifica è basata su metodologie di *keyword spotting* all'interno di segmenti conversazionali.

Se ricordiamo quanto abbiamo detto a proposito delle difficoltà di riconoscimento acustico causate dalla variabilità della produzione linguistica, possiamo facilmente immaginare come la possibilità di errori del paziente e i fenomeni tipici del parlato afasico,

⁹ 2009-2012 <https://vithea.l2f.inesc-id.pt>

come i tentativi, le disfluenze, le esitazioni e gli errori di tipo fonemico, debbano tutti essere presi in attenta considerazione nella predisposizione e nell'adattamento della tecnologia di riconoscimento. Nel progetto Vithea tale adattamento del sistema di riconoscimento acustico del portoghese è stato affrontato creando e utilizzando nel training del sistema un corpus di parlato afasico.

L'*Aphasia Portuguese Speech Corpus* è stato ottenuto registrando circa due ore e mezzo di parlato di ventisei soggetti con afasia all'interno di sedute di riabilitazione. La registrazione è avvenuta senza particolari accorgimenti per minimizzare le fonti di rumore ambientale e senza l'utilizzo di microfoni di alta qualità. Il segnale vocale registrato è stato successivamente sottoposto alle fasi canoniche di trattamento dei dati (segmentazione, trascrizione, annotazione). I dati linguistici annotati sono stati utilizzati per il training del riconoscitore acustico al fine di migliorare le sue prestazioni nel riconoscimento del parlato afasico.

Vithea, come altri progetti simili per la lingua inglese, sembra fornire una soluzione interessante perché il sistema consente al paziente di accedere autonomamente al sito di riabilitazione e di fare esercizio a casa propria, aumentando così l'opportunità di giovare di un trattamento intensivo. I terapeuti della Facoltà di Medicina hanno valutato che i soggetti sperimentali che hanno utilizzato il sistema hanno mostrato un guadagno di cinque punti nella scala AQ-WAB, oltre a miglioramenti nella capacità complessiva di comunicazione verbale e nell'abilità conversazionale. Per quanto riguarda il terapeuta, i vantaggi citati riguardano la facilità del monitoraggio a distanza, l'assistenza del sistema nella creazione di compiti personalizzati, oltre alla possibilità di impartire sedute di riabilitazione più frequenti.

Utilizzo di software di riconoscimento vocale per la riabilitazione dell'afasia per la lingua italiana

I colleghi portoghesi del progetto Vithea hanno valutato i risultati per un solo compito riabilitativo, quello relativo alla denominazione, sia pure con presentazione multimodale degli stimoli. In effetti, anche per quanto riguarda progetti analoghi, non esistono ancora dati sperimentali sufficienti sull'efficacia di questo approccio per *ampie popolazioni* di pazienti parlanti lingue diverse¹⁰.

Per quanto riguarda l'italiano, se da un lato abbiamo a disposizione sistemi commerciali e sperimentali di riconoscimento della voce con ottime prestazioni in termini di accuratezza per parlanti sani, non disponiamo ancora di raccolte di parlato di soggetti afasici che siano state trattate e rappresentate in modo da poter essere utilizzate per il training di tali

¹⁰ Wallesch & Johannsen-Horbach, Computers in aphasia therapy: Effects and Side-effects, *Aphasiology*, Vol 18, 3, 2004, pp. 223-228

sistemi. In altri termini non abbiamo un corpus di parlato spontaneo di soggetti con afasia appositamente raccolto all'interno di sedute di riabilitazione, contenente anche "metadati" essenziali quali il tipo di afasia, l'età dei soggetti, la valutazione neuropsicologica e altri dati linguistici pre-morbosi.

Un progetto di integrazione di tecnologie di riconoscimento della voce in sistemi di riabilitazione computerizzate per gli afasici di lingua italiana non può prescindere dalla creazione di un tale corpus né dall'individuazione, all'interno di un modello neuropsicologico, dei compiti riabilitativi più adatti alla somministrazione assistita dal computer.