

Computer-Assisted Language Learning e prosodia dell'italiano

Antonio Romano, Valentina De Iacovo, Marco Palena
LFSAG, Dip. di Lingue e L.S. e C.M. – Università di Torino

Abstract

Comparing utterances of learners of a foreign language (FL) with those of native speakers is receiving more attention thanks to the increasing number of applications arising in the field of computer-assisted teaching (Cazade, 1999; Delmonte, 1999) or computer-assisted and data-driven language learning (*CALL*; Levy, 1997; *DDL*; Corino, 2019). Since Jones (1909), a metalinguistic sense of intonation and general properties of the utterance may be considered useful for foreign language learning (James 1976, De Bot, 1983). Nevertheless, studies on prosodic inter-speaker variation brought to light a rhythmic-intonational variability that should not be reduced to a few acceptable patterns. Therefore, a broad knowledge of prosodic patterns should be part of the learner's linguistic background as much as their variation within a language system (Marotta & Sardelli, 2009; Henriksen 2013; De Iacovo 2019).

Nowadays, automatic segmentation systems and speech analysis tools make possible to compute meaningful parameters that characterise different

prosodic components and help the learner to develop his/her prosodic competence (Chun, 2013; Frost & Picavet, 2014, Romano & Giordano, 2017, De Iacovo & Romano, 2019). In fact *CALL* systems encourage L2 learners to grant more attention to acoustic features (f_0 , intensity, duration) and other time-related speech phenomena correlated to linguistic properties (number of syllables, lexical accent, speech speed, pauses; see Munro & Derwing, 1999).

In our project «*CALL-UniTO*» we carried out an experiment by using a Telegram chatbot to evaluate some prosodic features in the reading of sentences by Italian L2 learners by referring them to a sound database of 250 native speakers (*BDn*). The chatbot provides the learner with a pre-recorded set of questions and the answers are collected and compared to those of the best-matching utterance by a native speaker in the *BDn*. A score is assigned to each pair of sentences (Franco *et alii* 2010) on the basis of a distance measurement performed by a procedure developed and tested within the *AMPER* project (Moutinho *et alii*, 2011; Romano *et alii*, 2014):

f_0 values are extracted and the curves stylised, whereas a further weighting with duration patterns produces a series of prosodic vectors. An algorithm computes a correlation between vectors within the whole array and selects the best-matching sample. The user is then warned about his performance (with a three-value score) and is allowed to repeat his/her task if (s)he wishes to get closer to the more similar utterance by the selected native speaker.

1. Introduzione: motivazione e strategia

Numerose applicazioni hanno fatto progredire rapidamente negli scorsi decenni il campo della didattica assistita al computer (come mostrano, tra l'altro, vari contributi in Busà, 2012). Il confronto tra le produzioni di apprendenti di una lingua straniera e quelle di parlanti nativi con finalità didattiche (inaugurato da lavori come quelli di Cazade, 1999, e Delmonte, 1999, e proseguito con contributi come quello di Trouvain & Gut, 2007) sta trovando oggi nuova linfa nei filoni di studi *CALL* (*Computer-Assisted Language Learning*; sin da Levy, 1997) e *DDL* (*Data-Driven Language Learning*; cfr. ora Corino, 2019). Nel nostro caso, guardando alla consistenza ritmico-intonativa degli enunciati, riconosciamo in Jones (1909) un primo rilevante tentativo di usare metodi sperimentali per

aiutare gli utenti di una lingua (nativi o apprendenti di LS) a sviluppare un senso metalinguistico dell'intonazione e delle proprietà enunciative generali (James 1976, De Bot, 1983). Tuttavia, gli studi successivi sulla variazione prosodica all'interno della stessa comunità linguistica hanno mostrato come le variabili dell'enunciazione determinino una variabilità ritmico-intonativa che non si riduce facilmente a un insieme limitato di *pattern* (Marotta & Sardelli, 2009; Crocco, 2017). Questo ha suggerito la necessità di sviluppare una diffusa conoscenza delle principali dimensioni di variazione degli schemi intonativi più ricorrenti come parte del bagaglio di conoscenze di un apprendente di LS (Henriksen 2013; De Marco *et alia*, 2014; Romano & Giordano, 2017).

Per tali motivi, il lavoro presentato in quest'articolo si è orientato verso una soluzione che prevede l'estrazione di un certo numero di parametri che caratterizzano i principali aspetti prosodici di un enunciato e l'automatizzazione di una misura di calcolo dell'approssimazione con cui un apprendente si avvicina a produzioni *native-like* della sua LS.

Rifacendoci al concetto di competenza prosodica sviluppato da vari autori (Chun, 2013; Frost & Picavet, 2014, Romano & Giordano, 2017, De Iacovo & Romano, 2019), il nostro obiettivo è quello di sensibilizzare gli

apprendenti ai principali tratti prosodici (dipendenti da variabili acustiche come f_0 , intensità e durata) e ai fenomeni che incidono sul *timing* di un enunciato (velocità d'eloquio, pausa-zione etc.) in funzione dei vincoli linguistici (strutture sillabico-accentuali, gerarchie di prominenza etc., cfr. Munro & Derwing, 1999; De Meo & Pettorino, 2012).

Ci siamo dedicati a questo scopo nell'ambito di un progetto intitolato «CALL-UniTO – Per una didattica assistita nell'apprendimento dell'italiano parlato e nella valutazione linguistica delle conoscenze disciplinari», finanziato dalla Fondazione CRT come proseguimento del lavoro di ricerca «Dare voce ai corsi online – Per una didattica assistita della pronuncia e dell'intonazione nell'apprendimento dell'italiano L2»¹.

1. «Call-UniTO» ha beneficiato di un'erogazione ordinaria accordata al LFSAG (sotto la supervisione di Antonio Romano) e destinata a un assegno e una borsa di ricerca, conseguiti rispettivamente da Valentina De Iacovo e da Marco Palena, nell'arco del periodo dicembre 2020-novembre 2021. «Dare voce ai corsi online» aveva invece previsto un altro assegno di ricerca finanziato dalla stessa fondazione (erogazioni ordinarie 2019-20, responsabile del progetto Carla Marellò, Dip. di Lingue e L.S. e C.M. dell'Università degli Studi di Torino) e, nelle sue fasi di realizzazione tecnica, era stato affidato agli stessi A. Romano, V. De Iacovo e M. Palena, rispettivamente autori dei paragrafi 1., 2.1 e 2.2. del presente articolo. Mentre le parti comuni e il §4 si devono a tutti e tre gli autori, il §3 è interamente a cura di VD e MP.

In «CALL-UniTO» abbiamo svolto una serie di esperimenti usando un *chatbot Telegram*[®] che prevede una verifica di comprensione, impegnando l'apprendente anche su aspetti di cultura generale, saggiando le sue competenze conversazionali e candidandosi per la valutazione di conoscenze disciplinari². Il *chatbot* è inoltre in grado di valutare sommariamente alcune caratteristiche prosodiche nella lettura di frasi in italiano confrontando le produzioni dell'apprendente con quelle di una base di dati di produzioni linguistiche simili da parte di un campione di parlanti nativi (v. §2.1)³.

Il *chatbot* fornisce all'utente, in forma scritta e parlata, un insieme pre-registrato di domande e raccoglie le sue risposte confrontandole con quelle del parlante nativo con caratteristiche vocali e intonative più simili alle sue. In seguito a una segmentazione automatica operata grazie a un servizio *WebMAUS* (Schiel, 1999; Kisler *et alii*, 2017), l'algoritmo di estrazione

2. L'attività di ricerca ha permesso di coinvolgere un numeroso gruppo di studenti, creando una collaborazione con il Politecnico di Torino, partner fondamentale per la costituzione del *bot* e l'estrazione dei dati. I nostri ringraziamenti si rivolgono a tutti i partecipanti al progetto e a Luciano Lavagno (DET) e Giampiero Cabodi (DAUIN).

3. Come si mostrerà, alla base iniziale di 250 parlanti si sono aggiunti altri 150 parlanti nativi che hanno aiutato a testare le prestazioni del bot.

dei valori f_0 lavora con uno *script* per PRAAT messo a punto per il progetto AMPER (Romano *et alii*, 2014) e con sistemi di stilizzazione e valutazione delle distanze testati ormai su diverse lingue (Moutinho *et alii*, 2011). Le curve di f_0 stilizzate sono dapprima sottoposte a una ponderazione in base ai *pattern* di durata definendo un vettore per ogni enunciato che, insieme alla registrazione originale, arricchisce la base di dati degli utenti (BDI). Alle coppie ottenute associando il vettore in *input* ai vettori corrispondenti della BDI è poi assegnato un valore di distanza (o di prossimità) secondo una procedura disposta in altri studi (Franco *et alii*, 2010). Sulla base della selezione della coppia di vettori con valore più alto di prossimità si attribuisce poi un punteggio alla singola produzione in *input* (De Iacovo & Palena e De Iacovo *et alii*, in c. di p.; v. §2.2). Le difficoltà e le soluzioni adottate per aumentare l'affidabilità delle valutazioni sono dettagliate invece al §3, insieme a una descrizione delle modalità con cui si attribuisce un punteggio finale complessivo dell'intera prestazione.

Al di là degli *score* risultanti, l'utente è comunque gratificato per l'incremento che indirettamente riceve nelle sue attività di riflessione metaprosodica (Trouvain & Gut, 2007): come si vedrà, gli apprendenti sperimentano, infatti, una varietà di soluzioni enunciate che sviluppano l'attenzione nei

riguardi di abilità comunicative necessarie su diversi piani (conversazionale, sintattico, informativo ed emotivo) per i quali possono poi cercare indicazioni più dettagliate (Canepari, 1985, 2012).

2. Caratteristiche generali della BDI e dell'accesso ai dati

2.1. Consistenza del corpus di riferimento

Le fasi preliminari del progetto si erano concretizzate in una serie di attività didattiche *online*, riservate agli studenti dell'Università di Torino. Tra quelle di aree linguistica, alcune preludevano a una presa d'atto delle qualità linguistiche dell'intonazione, con l'obiettivo di valutare la loro importanza nello sviluppo della competenza orale in LS (principalmente pronuncia e aspetti sovrasegmentali nella produzione di brevi enunciati in lettura). Le attività glottodidattiche predisposte erano strutturate in percorsi che proponevano l'utilizzo di strumenti digitali rivolti alla valutazione dell'italiano L2 parlato da studenti non italo-foni. Nella parte iniziale di questo progetto si è potuto procedere poi a una prima valutazione dei dati di studenti stranieri grazie alla definizione di un corpus di riferimento di italo-foni (consolidato nei mesi grazie a nuove registrazioni). La valutazione, basata sul confronto intonativo qui descritta, ha permesso di far emergere nuovi interessanti spunti di indagine legati proprio alla complessità del corpus orale analizza-

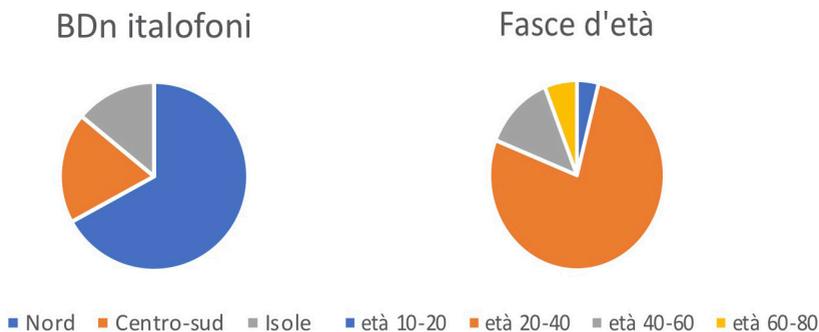


Fig. 1. Composizione dell'archivio di dati raccolti nella prima fase (parlanti nativi, *BDn*).

to. I dati (degli italofoeni e non) sono ancora attualmente valutati e classificati in termini percettivi su diversi piani (naturalità, qualità generale delle registrazioni), ma una prima valutazione degli esperti coinvolti nel progetto ha permesso una comparazione tra i valori acustici e percettivi che restituì delle soglie di accettabilità (sono state valutate specificatamente le componenti ritmico-intonative, la velocità d'eloquio e le pause).

Nello stesso tempo, in questa fase, è avvenuto lo sviluppo del *chatbot*, ottimizzando le modalità di interazione con l'utente e provvedendo a una progressiva raccolta di dati orali preorganizzati nella base di dati di riferimento (*BDn*). La *BDn* include, al momento, le registrazioni di 400 parlanti madrelingua (300 F, 100 M) che producono un set di dieci frasi complesse costruite in modo da dare un'adeguata rappresentazione di alcune soluzioni intonative di un parlato informale e formale su contenuti di natura tecnico-scientifica.

Per testare anche le abilità conversazionali degli apprendenti, include anche alcune espressioni quotidiane con più strutture intonative (dichiarativa, interrogativa, continuativa-elencativa).

L'archivio di dati raccolti può essere sommariamente descritto per mezzo dei grafici di Fig. 1.

Come si vede, attualmente la *BDn* comprende quindi pronunce di parlanti di varia origine regionale da Nord (67%) a Centro-Sud (19%) e isole (14%), v. Fig. 1a, e di varie fasce d'età (giovanissimi, giovani, adulti, anziani), v. Fig. 1b⁴.

4. La *BDn* offre anche una considerevole quantità di dati che si propone per uno studio sulle caratteristiche vocali di un campione di parlanti nativi di italiano (400!) che lascia prospettare interessanti ricadute sulla conoscenza dei *range* di variazione di qualità linguistiche e fisiologiche. I dati raccolti consentono di produrre nell'immediato documentazione originale riferibile a una popolazione per la quale non si dispone ancora di dati definiti (v. Romano & De Iacovo, in questo numero) e che offrirà un riferimento per valutazioni future sulla qualità della voce.

2.2. *Struttura del chatbot*

L'adozione di un *chatbot* come strumento di didattica assistita presenta diversi vantaggi (cf. Lezhenin *et alii*, 2017; Fernoagà *et alii*, 2018). Innanzitutto prevede una modalità di interazione che, per quanto orientata a valutazioni sul piano dell'oralità, è facilitata da modalità d'interazione che prevedono anche un testo scritto. Ciò rende questi strumenti più facilmente fruibili rispetto ad applicazioni dedicate o ai cosiddetti *Learning Management System* (LMS) come *Moodle* o *Blackboard*, il cui maggior ostacolo alla fruibilità è rappresentato proprio dalla necessità da parte degli utenti di adattarsi a un'interfaccia poco familiare. Un *chatbot* permette inoltre un'interazione con gli apprendenti in maniera strutturata e asincrona: strutturata perché organizzata secondo schemi predefiniti,

volti a ridurre le ambiguità; asincrona perché non vincolata a intervalli di tempo continuativi, ma fruibile liberamente secondo le proprie disponibilità di tempo. Una volta individuata la risposta corretta, l'apprendente invia la registrazione della frase letta sotto forma di messaggio vocale. L'enunciato ricevuto viene così processato automaticamente dal *bot* al fine di ottenere una valutazione del livello intonativo del parlante. Il primo passo nella catena di processamento ha un duplice scopo: garantire un livello di qualità minimo per il segnale audio di *input* e preparare quest'ultimo alle elaborazioni successive. L'audio viene innanzitutto convertito in formato *.wav* a singolo canale, ricampionato (se necessario) a 48 kHz e ripulito dal rumore di fondo costante. Vengono poi stimati il valore medio di ampiezza e il rapporto

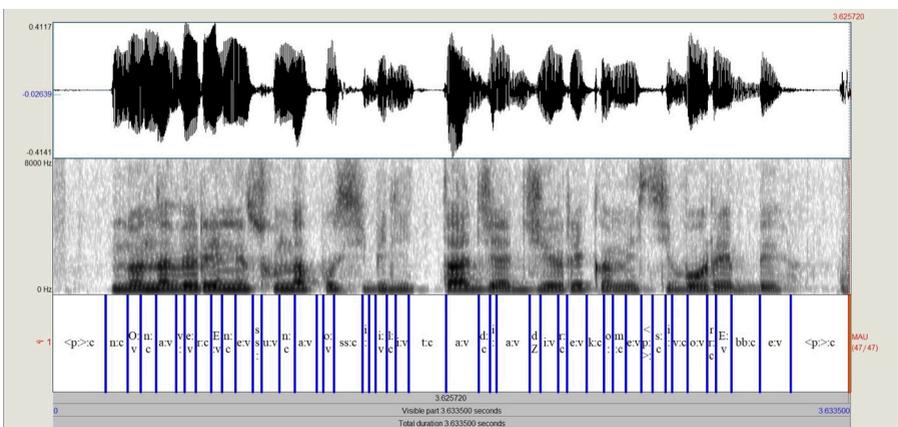


Fig. 2. Esempio di segmentazione dell'enunciato «Non avere nessuna possibilità di agire come si vorrebbe.» da parte di un parlante nativo della BDn (n. §2.1).



Fig. 3. Esempio di allineamento tra enunciati con segmenti diversi (enunciati «Frasi di prova» e «Frasi prove»).

segnale/rumore dell'audio risultante. Nel caso in cui i valori stimati risultino inferiori a determinate soglie predefinite, il *chatbot* richiede all'apprendente di registrare un nuovo enunciato in un ambiente meno rumoroso e/o parlando a un tono di voce più alto. Il segnale vocale viene successivamente sottoposto a segmentazione mediante il servizio *WebMAUS Basic* (Schiel, 1999; Kisler *et alii*, 2017). Tale servizio, una volta inviato il segnale vocale come *input* e la trascrizione ortografica dell'enunciato fornita dal *bot*, ne restituisce una segmentazione in parole e segmenti sonori rispondenti alle attese fonologiche. La segmentazione fonetica dell'enunciato, fornita in formato *.TextGrid* per agevolare la successiva elaborazione mediante il software Praat, viene poi processata etichettando i singoli segmenti come vocali o consonanti (v. Fig. 2).

Uno script Praat viene poi richiamato per estrarre i valori di frequenza fondamentale (f_0) dei segmenti vocalici precedentemente identificati, ottenendo così un campione stilizzabile della

curva di altezza melodica che caratterizza l'enunciato⁵. L'analisi dell'intonazione avviene mediante il confronto tra la curva intonativa dell'enunciato dell'apprendente e i tracciati f_0 dei corrispettivi enunciati di parlanti madrelingua (v. §2.1), raccolti e valutati utilizzando la procedura automatica di cui sopra. Il confronto avviene mediante il calcolo di una misura di correlazione che confronta per ogni frase tre punti di f_0 (iniziale, centrale, finale) di ogni segmento vocalico individuato da MAUS. Data la natura estemporanea dei segnali vocali registrati e la modalità automatica della procedura

5. Sin dal 2005, alcuni *script* per PRAAT, parzialmente compatibili con le esigenze del progetto AMPER, sono stati implementati da P. Barbosa (Campinas, Brasile). Dal 2008, uno di questi – adattato da A. Rilliard (LIMSI, Orsay-Parigi) – è stato diffuso in una versione (AMPER_PRAAT_Textgrid2Txt.psc) in grado di convertire una segmentazione / etichettatura eseguita con PRAAT in un file .txt che riassume le informazioni relative al *timing* e alle variazioni di f_0 e intensità in ciascuno dei segmenti che caratterizzano le sillabe dell'enunciato (Romano *et alii*, 2014).

di segmentazione, i segmenti vocalici rilevati per gli enunciati degli apprendenti e quelli rilevati per i corrispettivi enunciati di parlanti madrelingua si presentano in molti casi piuttosto differenti. Pertanto, prima di calcolare la correlazione, i segmenti dei due parlanti vengono allineati sulla base sia dell'informazione fonetica in essi contenuta sia della loro posizione temporale. Questa procedura di allineamento serve per mettere in corrispondenza i segmenti dei due parlanti in modo da sapere quali confrontare tra loro per il calcolo della correlazione, quindi solo i segmenti per cui si è trovata corrispondenza sono inclusi nel calcolo. Sono esclusi invece tutti i segmenti vocalici per cui non si trova corrispondenza fonetica tra i due enunciati. Ad esempio, se l'enunciato di un parlante include in una certa posizione un segmento etichettato come [i] (v. Fig. 3) mentre l'enunciato dell'altro parlante non lo include, quel segmento non viene preso in considerazione come punto ai fini del calcolo della correlazione. Questo perché per calcolare la correlazione sul modello di Moutinho *et alii* (2011) si ha bisogno di due serie di dati della stessa numerosità⁶.

6. Oltre ai fenomeni di riduzione/cancellazione di segmenti sonori in parlanti nativi che ipoarticolano alcuni passaggi, un caso molto frequente nelle prove degli apprendenti LS è quello dell'aggiunta di segmenti vocalici in false partenze o esitazioni nelle pause (v. §4; cfr. Savy, 2006).

3. Funzionalità del *chatbot* e svolgimento del *task*

Il *chatbot* prevede l'interazione con l'utente attraverso domande e risposte basate su una serie di domande a risposta chiusa (quiz) che possono avere carattere generale di comprensione linguistica o essere legate a un particolare ambito disciplinare (ad esempio tecnico-scientifico)⁷. Per svolgere il *task* è necessario aver installata l'applicazione Telegram® su un dispositivo *smart phone*. La *chat* si avvia digitando «CallUnitoDevbot» e, una volta digitato /*start*, si viene indirizzati (tramite un link esterno) alla registrazione utente che permette di inserire alcune informazioni di carattere sociolinguistico (lingua materna, livello del *CEFR* in italiano, paese/città d'Italia in cui si è eventualmente vissuto, Fig. 4).

Una volta effettuata l'iscrizione, l'utente viene indirizzato ad alcune attività preliminari: la produzione di parlato letto e spontaneo (così da avere un ulteriore riferimento in produzione orale per l'analisi successiva). A questo punto è possibile cominciare il *task*, seguendo le indicazioni date progressivamente dal *bot*: nelle istruzioni iniziali si legge che sulla base della lettura di 10 risposte, l'utente avrà

7. Un utile modello di costruzione di quiz conversazionali autoguidati è offerto da Pereira (2016).

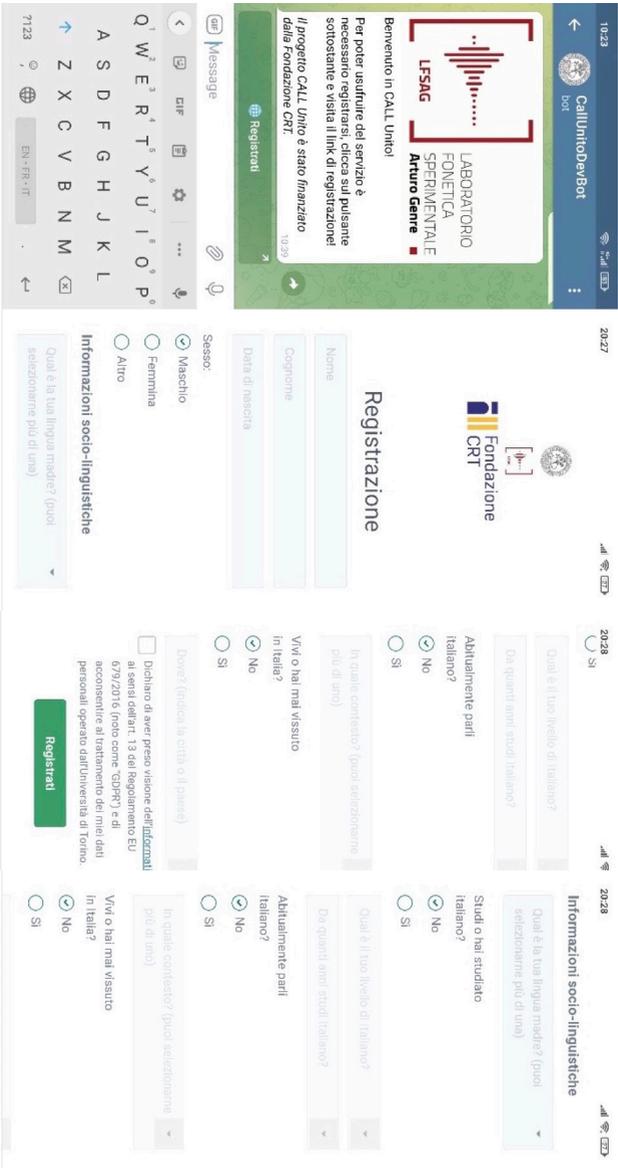


Fig. 4. Schermate di registrazione al chatbot CALL-Unito.

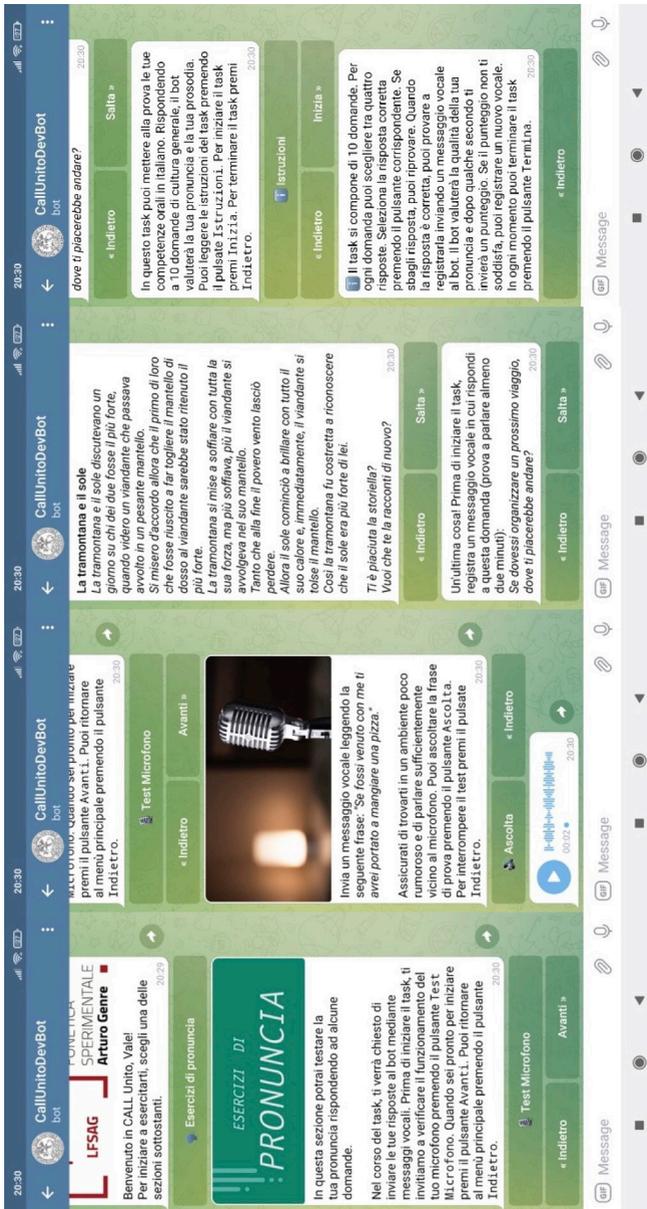


Fig. 5. Schermate di svolgimento di attività preliminari col chatbot CALL-UnitO (prova microfono, prova lettura etc.).

un punteggio indicativo della propria pronuncia e intonazione (v. Fig. 5).

Dopo aver risposto correttamente a una domanda, il *chatbot* mostra la risposta corretta con un altro messaggio e richiede di registrare un messaggio vocale con la risposta corretta. Fatto ciò, bisogna attendere qualche secondo per avere un punteggio individuale. Il punteggio è stabilito sulla base della distanza prosodica (v. §2.2): la frase prodotta viene infatti comparata istantaneamente con quelle presenti nella *BDn* per trovare quella che più si avvicina in termini intonativi. A questo punto il *chatbot* restituisce un punteggio indi-

cativo (verde=buono, giallo=medio, rosso=scarso, v. Fig. 6) corrispondente alla percentuale di correlazione stabilita tra i due enunciati più simili sul piano intonativo (verde >65%, giallo >40%).

Se la risposta data è sbagliata, viene segnalato e l'utente può riprovare. In questa fase non è prevista nessuna penalizzazione, non si hanno quindi limiti di tentativi per le risposte. Compilate le 10 domande, l'utente riceve alcune informazioni riassuntive della propria *performance* orale (livello intonativo medio, risposte corrette, tentativi totali, frasi registrate, v. Fig. 7).

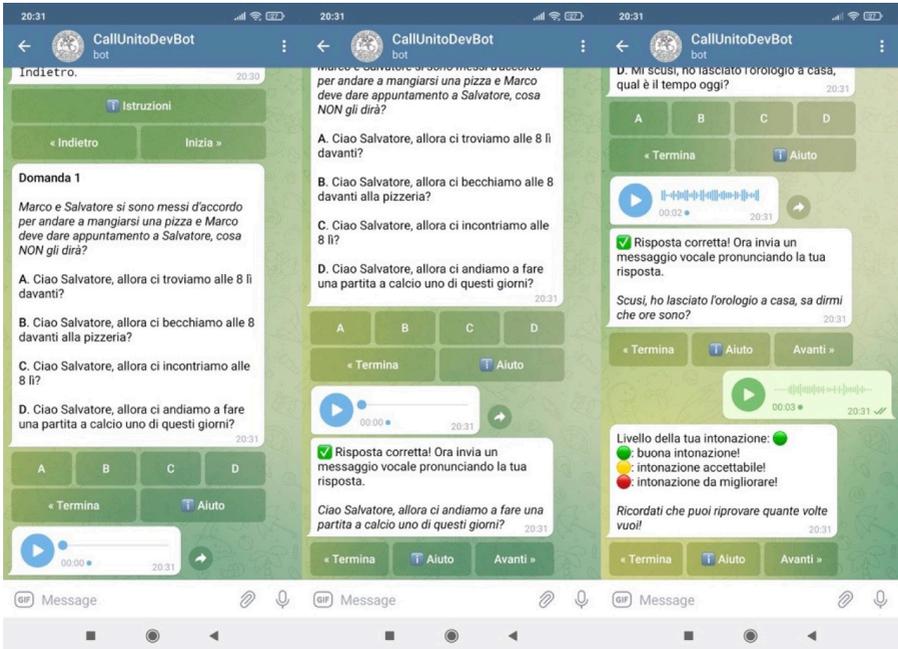


Fig. 6. Schermate di interazione per la raccolta di enunciati valutabili e restituzione di valore di prossimità con enunciati comparabili prodotti dai parlanti della *BDn*.

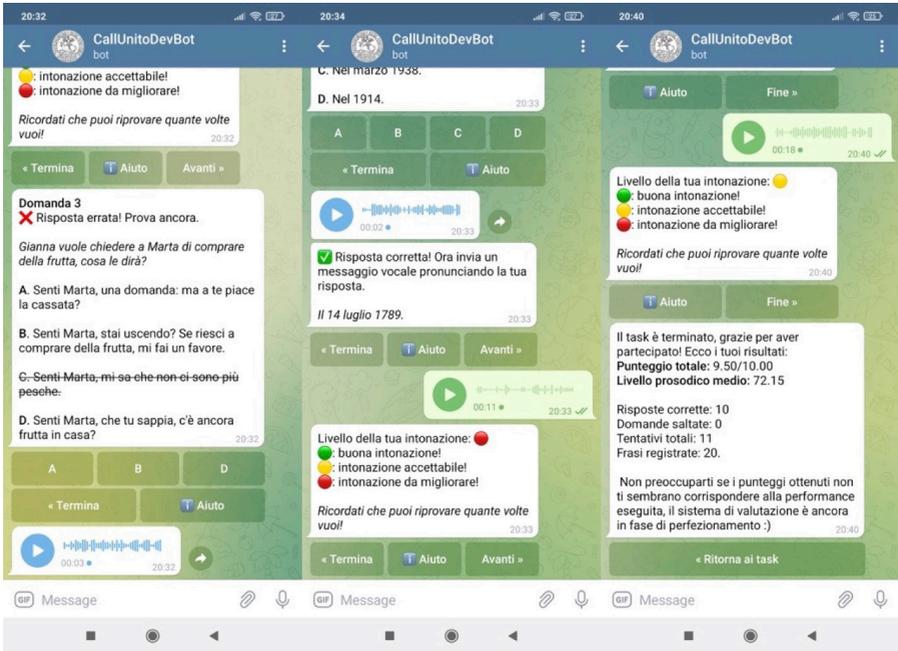


Fig. 7. Schermate di interazione nel caso di risposte errate o di livelli di correlazione bassi. Schermata finale con bilancio dell'attività complessiva.

4. Risultati e sviluppi futuri

Il dispositivo è stato testato su un campione di 60 apprendenti di italiano LS, in alcuni casi classificati come *native-like*, che hanno contribuito alla creazione di un primo campione (*BDI*). I partecipanti all'esperienza hanno messo alla prova le proprie prestazioni enunciative rispetto a quelle dei parlanti madrelingua, ritrovandosi in una situazione in cui anche solo l'effetto di un'autovalutazione ha portato sicuri benefici in un ambito di produzione orale sul quale raramente l'apprendente (e il docente di lingua) si sofferma.

La creazione di una *BDI* così strut-

turata consente ora di lavorare alla definizione di modalità di segnalazione delle differenze e alla ricerca di metodi correttivi più espliciti oltre che di metodi di classificazione automatica più affidabili (Papi *et alii*, 2020).

Allo stato attuale i dati sono stati annotati da tre diversi operatori esperti in base alla maggiore o minore accettabilità intonativa (su tre gradi di spontaneità, indicati con 1÷3, v. Tab. D). Le stringhe di caratterizzazione così definite si prestano per addestrare un algoritmo in grado di sostituire l'operatore umano lungo una linea che distingue un parlato artefatto o impac-

val_pair Ut_ref	% occorrenza	avg_corr	std_dev_corr
1-1	0,24%	61%	0
1-2	0%	0%	0
1-3	3,17%	63%	12,19
2-1	0,24%	57%	0
2-2	2,68%	60%	14,29
2-3	23,17%	70%	12,27
3-1	0,49%	60%	22,26
3-2	3,66%	74%	14,25
3-3	66,34%	73%	13,47

Tab. I. Occorrenza di corrispondenze tra livelli di qualità dell'enunciato dell'utente (Ut) vs. enunciato più vicino nella BDn (ref) e correlazione media (avg_corr) tra valutazione automatica e giudizi di esperti.

ciato (lettura istrionica o difficoltosa) da uno sciolto e spontaneo, sulla base di diversi parametri acustici (numero di sillabe, pause, velocità d'eloquio, durate parziali e complessive).

Circa il 90% dei dati della *BDI* finora esaminati ha presentato coerenza con dati di parlato non connotato; gli enunciati degli apprendenti hanno inoltre ricevuto una valutazione percettiva complessivamente positiva (alta o media) nel 96% di casi. Gli utenti analizzati finora (60) provengono da diverse parti d'Europa (Francia, Spagna, Inghilterra, Polonia, Portogallo, Romania), Russia, Nord-Africa (Marocco) e Sud-America (Brasile); essi hanno dichiarato di possedere un livello generalmente medio-alto in italiano (B2-C1), il che concorda con la valutazione percettiva della loro prestazione intonativa che è generalmente alta (70%) o media (26%) (Tab. I).

Confrontando le valutazioni assegnate a ciascuna coppia di enunciati (dell'utente e dell'esempio più vicino nella *BDn*, 1ª colonna in Tab. I) si riscontra una netta prevalenza di coppie in cui entrambi i modelli di pronuncia sono stati giudicati buoni (3-3), mentre alcune associazioni sono risultate trascurabili (es. 1-1, 1-2, 2-1, 3-1). D'altra parte, valutando la coerenza di giudizi dati dal sistema automatico e dagli operatori umani (2ª col.) si verifica una correlazione media generalmente alta (57÷74%) all'interno di tutte le combinazioni, con oscillazioni praticamente nulle in diversi casi poco rappresentati (di nuovo 1-1, 1-2, 2-1)⁸.

Si noti la frequenza di casi in cui il punteggio attribuito all'associazione

8. Al contrario nei casi più improbabili – come 3-1 – la percentuale di accordo è più variabile (a una correlazione del 60% corrisponde una deviazione standard di 22 punti circa).

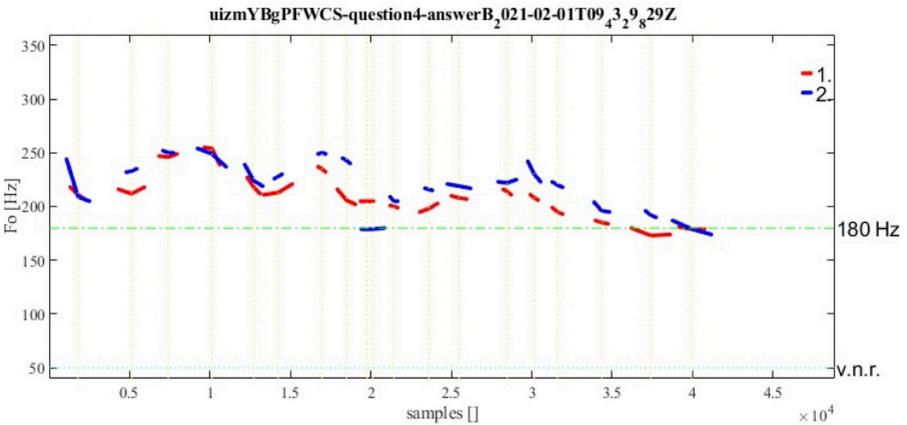


Fig. 8. Confronto tra profili melodici secondo il modello AMPER (curve di f_0 con segmenti allineati). Caso di buona approssimazione dell'enunciato dell'apprendente (blu) a quello di un parlante madrelingua della BDn (rosso): «Il quattordici luglio millesettecentottantanove».

sia migliorativo (2-3): nel 23% dei casi, infatti, una risposta data dall'utente è stata associata dal *chatbot* a un enunciato presente nella *BDn* con una valutazione più alta. Questo porta a riflettere sulla complessità delle componenti prosodiche che dovrebbero essere analizzate separatamente: la distanza prosodica (in casi di curve simili, es. in Fig. 8) ha infatti portato all'associazione di prove non eccellenti con enunciati modello ritenuti ottimali. Tener conto delle modalità di realizzazione di accenti e profili intonativi non è quindi sufficiente⁹: altri importanti parametri meritano di essere analizzati in termini quantitativi e qualita-

tivi; tra questi, oltre alle le pause e alle modalità di accentazione, sembrano rilevanti soprattutto le durate parziali e totali degli enunciati e le variazioni locali nella velocità d'eloquio.

È, inoltre, in corso una fase di quantificazione e soluzione dei casi in cui gli enunciati si presentino difforni da quelli attesi a causa di un ordine di parole diverso e/o della presenza di frammenti involontari di parlato (babbettii, esitazioni, false partenze) che causano differenze tali da impedire il conseguimento di una minima distanza prosodica. Una possibilità adottata per monitorare la stringa sillabica elicitata dall'utente consiste nel limitare entro un certo numero le sillabe possibili in ciascun enunciato così da garantire una soglia di sbarramento oltre la quale riproporre la lettura.

9. In questo caso la ponderazione non ha fatto pesare abbastanza le differenze sui rapporti di durata ai confini delle unità tonali.

Tra gli sviluppi futuri, vorremmo testare infatti enunciati di più difficile lettura, che prevedano la presenza di semplici formule matematiche, sigle, acronimi e forestierismi che costellano la nostra stampa quotidiana generalista e, più in particolare, le notizie a sfondo economico-sociale, senza trascurare le espressioni richieste dalle soluzioni enunciative tipiche di alcuni linguaggi specialistici. Su questi aspetti nessuna formazione specifica è prevista nei curricoli di avvicinamento alle materie tecnico-scientifiche e nessuna indagine nazionale dà ancora indicazioni esaustive.

Riferimenti bibliografici

- BOERSMA P. & WEENINK D. (2018). «Praat: doing phonetics by computer [Computer program]». Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>
- BUSÀ M.G. (2012). «The role of prosody in pronunciation teaching: a growing appreciation». In M.G. BUSÀ & A. STELLA (eds.), *Methodological Perspectives on Second Language Prosody*, Padova: CLEUP, 101-105.
- CANEPAI L. (1985). *L'intonazione. Linguistica e paralinguistica*. Napoli: Liguori.
- CANEPAI L. (2012). *Italian intonation: neutral & regional accents* (pdf file retrieved from <https://canipa.net/doku.php?id=en:pdf>).
- CAZADE A. (1999). «De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues». In *Apprentissage des Langues et Systèmes d'Information et de Communication*, 2(2), 3-32.
- CORINO E. (2019). *Data-Driven Learning: La linguistica dei corpora al servizio della didattica delle lingue straniere e del CLIL*, No. monografico di EL.LE, 8 (2) (doi.org/10.30687/ELLE/2280-6792/2019/02).
- CHUN D.M. (1998). «Signal analysis software for teaching discourse intonation». *Language Learning & Technology*, 2, 6177.
- CHUN D.M. (2002). *Discourse Intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.
- CHUN D.M. (2013a). «Computer-Assisted Pronunciation Teaching». In C. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, Oxford: Wiley-Blackwell (doi.org/10.1002/9781405198431.wbeal0172).
- CHUN D.M. (2013b). «Technology and Discourse Intonation». In C. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, Oxford: Wiley-Blackwell (doi.org/10.1002/9781405198431.wbeal1180).
- CROCCO C. (2017). «Everyone has an accent: standard Italian and regional pronunciation». In M. CERRUTI, C. CROCCO & S. MARZO (eds.), *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*, Berlin/New York: Mouton de Gruyter, 89-117.

DE BOT K. (1983). «Visual feedback of intonation I: Effectiveness and Induced Practice Behavior». *Language and Speech*, 26(4), 331-350.

DE IACOVO V. (2019). *Intonation Analysis on Some Samples of Italian Dialects: an Instrumental Approach*, Alessandria, Dell'Orso.

DE IACOVO V. & PALENA M. (in c. di p.). «La variazione prosodica in italiano: l'utilizzo di un chatbot Telegram per la didattica assistita per apprendenti di italiano L2 e nella valutazione linguistica delle conoscenze disciplinari». In *Studi AITLA*, Milano: Officinaventuno, in c. di p.

DE IACOVO V. & ROMANO A. (2019). «Data-Driven intonation teaching: an overview and new perspectives». *ELLE*, 8(2), 393-408.

DE IACOVO V., PALENA M. & ROMANO A. (in c. di p.). Evaluating prosodic cues in Italian: the use of a Telegram chatbot as a CALL tool for Italian L2 learners. In *Atti del convegno AISV di Zurigo 2020*, Milano: Officinaventuno, in c. di p.

DE MARCO A., SORIANELLO P. & MASCHERPA E. (2014). «L'acquisizione dei profili intonativi in apprendenti di italiano L2 attraverso un'unità di apprendimento in modalità blended learning». In A. DE MEO & M. D'AGOSTINO (eds.), *Varietà dei contesti di apprendimento linguistico, Studi AitLA 1*, 189-210.

DE MEO A. & PETTORINO M. (eds.), (2012). *Prosodic and Rhythmic Aspects*

of L2 Acquisition: The Case of Italian. Newcastle Upon Tyne: Cambridge Scholars Publishing.

DELMONTE R. (2009). «Prosodic tools for language learning». *International Journal of Speech Technology*, 12(4), 161-184.

FERNOAGĂ V., STELEA G.A., GAVRILĂ C. & SANDU (2018). «Intelligent education assistant powered by chatbots». In *The International Scientific Conference eLearning and Software for Education*, 2, 376-383.

FRANCO H., BRATT H., ROSSIER R. RAO GADDE V., SHRIBERG E., ABRASH V. & PRECODA K. (2010). «Eduspeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications». *Language Testing*, 27(3), 401-418.

FROST D. & PICAUVET F. (2014). «Putting prosody first – Some practical solutions to a perennial problem: The Innovalangues Project». *Research in Language*, 12(3), 233-243.

HENRIKSEN N. (2013). «Style, prosodic variation, and the social meaning of intonation». *Journal of the International Phonetic Association*, 43(2), 153-193.

JAMES E.F. (1976). «The acquisition of prosodic features of speech using a speech visualizer». *International Review of Applied Linguistics in Language Teaching*, 14(3), 227-244.

JONES D. (1909). *Intonation curves. A collection of phonetic texts, in which intonation is marked throughout by means of*

curved lines on a musical staff. Leipzig-Berlin: Teubner.

KISLER T., REICHEL U. & SCHIEL F. (2017). «Multilingual processing of speech via web services». *Computer Speech & Language*, 45, 326-347.

LEVY M. (1997). *CALL: context and conceptualisation*. Oxford: Oxford University Press.

LEZHENIN Y., LAMTEV A., DYACHKOV V. BOITSOVA E., VYLEGZHANINA K. & BOGACH N. (2017). «Study intonation: mobile environment for prosody teaching». In *Proc. of IEEE International Conference on Cybernetics (CYBCONF)*, 3, 1-2.

MAROTTA G. & SARDELLI E. (2009). «Prosodiatopia: parametri prosodici per un modello di riconoscimento diatopico». In G. FERRARI, R. BENATTI, M. MOSCA (a cura di), *Linguistica e modelli tecnologici di ricerca* (Atti del XL Congresso Int. della SLI), Roma: Bulzoni, 411-435.

MOUTINHO DE CASTRO L., COIMBRA R., RILLIARD A. & ROMANO (2011). «Mesure de la variation prosodique diatopique en portugais européen». *Estudios de fonética experimental*, 20, 33-55.

MUNRO M.J. & DERWING T.M. (1999). «Foreign accent, comprehensibility, and intelligibility in the speech of second language learners». *Language Learning*, 49, 285-310.

PAPI S., TRENTIN E., GRETTNER R., MATASSONI M. & FALAVIGNA D. (2020). «Mixtures of deep neural experts for

automated speech scoring». In *Proceedings of Interspeech 2020*, 3845-3849.

PEREIRA J. (2016). «Leveraging chatbots to improve self-guided learning through conversational quizzes». In *Proceedings of the fourth international conference on technological ecosystems for enhancing multiculturality*, TEEM '16, ACM Press, New York: 911-918.

ROMANO A. & GIORDANO G. (2017). «Esperienze e riflessioni sulla didattica assistita dell'intonazione in italiano, inglese e francese». In A. DAMASCELLI (ed.), *Digital Resources, Creativity, Innovative Methodologies and Plurilingualism: New Approaches to Language Teaching and Learning*, Newcastle upon Tyne: Cambridge Scholars Publishing, 176-200.

ROMANO A. & DE IACOVO V. (2021). «Statistiche di f_0 per 200 parlanti di italiano». In questo numero.

ROMANO A., CONTINI M. & LAI J.P. (2014). «L'Atlas Multimédia Prosodique de l'Espace Roman: uno strumento per lo studio della variazione geoprosodica». In F. TOSQUES (ed.), *20 Jahre digitale Sprachgeographie*, Berlin: Humboldt-Universität-Institut für Romanistik, 27-51.

SAVY R. (2006). «Specifiche per la trascrizione ortografica annotata dei testi raccolti». In F. ALBANO LEONI & R. GIORDANO (a cura di), *Italiano Parlato. Analisi di un dialogo*, Napoli: Liguori, 1-37.

SCHIEL F. (1999). «Automatic phonetic transcription of non-prompted speech».

In *Proceedings of the International Conference of Phonetic Sciences 1999*, 607-610.

TROUVAIN J. & GUT U. (2007). *Non-native prosody. Phonetic description and teaching practice*. Berlin: Mouton de Gruyter.