

TEST DI INTELLEGIBILITÀ DEL PARLATO PERSONALIZZATI IN REALTÀ VIRTUALE: DEFINIZIONE
DI SCENARI AUDIOVISIVI E SVILUPPO DI UNO STRUMENTO CLINICO

Original

TEST DI INTELLEGIBILITÀ DEL PARLATO PERSONALIZZATI IN REALTÀ VIRTUALE: DEFINIZIONE DI SCENARI
AUDIOVISIVI E SVILUPPO DI UNO STRUMENTO CLINICO / Rosso, R.G., Urbanelli, A., Guastamacchia, A.. - (2025).
(51° Convegno Nazionale AIA Treviso 4-6 giugno 2025).

Availability:

This version is available at: 11583/3002458 since: 2025-08-18T18:26:27Z

Publisher:

Associazione Italiana di Acustica (AIA)

Published

DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in
the repository

Publisher copyright

(Article begins on next page)

TEST DI INTELLEGIBILITÀ DEL PARLATO PERSONALIZZATI IN REALTÀ VIRTUALE: DEFINIZIONE DI SCENARI AUDIOVISIVI E SVILUPPO DI UNO STRUMENTO CLINICO

Riccardo Giovanni Rosso (1), Anastasia Urbanelli (2), Angela Guastamacchia (3)

1) Dipartimento Energia – Politecnico di Torino, riccardo.rosso@polito.it

2) Dipartimento di Scienze Chirurgiche – Università degli Studi di Torino, anastasia.urbanelli@polito.it

3) Dipartimento Energia – Politecnico di Torino, angela.guastamacchia@polito.it

SOMMARIO

Lo sviluppo di test di intellegibilità basati su scenari realistici è fondamentale per valutare la perdita uditiva in modo aderente al deficit percepito. La realtà virtuale audiovisiva può essere utilizzata per ricreare scenari uditivi complessi, ma ulteriori studi sono richiesti per strutturarne le caratteristiche in ambito clinico e permettere una selezione mirata di questi alle necessità del paziente. Si propone una metodologia per definire tali scenari, corredata da un software per eseguire test di ascolto che permetta la selezione e configurazione degli scenari in modo adeguato, semplice e flessibile da parte del personale sanitario.

1. Introduzione

Nel contesto della valutazione dell'intelligibilità del parlato in ambito clinico, specialmente nel caso di procedure di adattamento degli apparecchi acustici, è essenziale utilizzare test che riflettano le condizioni reali di ascolto in cui gli individui interagiscono quotidianamente [1]. A tal proposito, la realtà virtuale può essere utilizzata per riprodurre scenari audiovisivi complessi all'interno dei quali effettuare test clinici [5]. Tuttavia, nonostante alcuni progressi, manca ancora uno standard programmatico consolidato per definire e utilizzare tali scene in modo uniforme per i test clinici. L'assenza di linee guida standardizzate rende difficoltoso garantire la coerenza, la selezione, la riproducibilità e l'affidabilità dei test in ambienti complessi. Sebbene iniziative come il framework CoSS (Common Sound Scenarios) [2] abbiano contribuito a classificare gli ambienti di ascolto più frequenti e rilevanti per gli utenti di apparecchi acustici, essi si concentrano principalmente sulla categorizzazione dei contesti sonori in base all'intenzione dell'ascoltatore e al tipo di compito piuttosto che sulla definizione delle scene da un punto di vista operativo e procedurale. Studi successivi, come quello di Wagener et al. [4], hanno esplorato la variabilità degli scenari uditivi vissuti dagli utenti di apparecchi acustici, evidenziando come la percezione dei benefici vari in base al contesto, con miglioramenti significativi in situazioni più semplici, ma limitati in ambienti complessi. Lo studio di Wu et al. [3], attraverso registrazioni quotidiane e analisi in situ, ha ulteriormente dettagliato questi scenari, ma anche in questo caso, la riproducibilità di tali ambienti secondo un pattern sistematico in un contesto clinico rimane un aspetto irrisolto.

Questo studio propone una metodologia per la definizione di scenari uditivi, facilitando, attraverso la tassonomia sviluppata, la selezione di condizioni d'ascolto realistiche per i test di intelligibilità. Inoltre, presenta un software progettato per supportare il personale clinico nella gestione e riproduzione dei test, consentendo di selezionare e configurare gli scenari audiovisivi in modo strutturato e modulare.

2. Metodologia per la definizione degli scenari audiovisivi

Per semplificare la raccolta delle scene audiovisive e fornire ai clinici uno strumento per testare l'intelligibilità in scenari più aderenti alla realtà, è stata definita una tassonomia che scompone ogni scena uditiva in quattro elementi principali, a loro volta descritti tramite un sottoinsieme di attributi:

1. *L'ambiente*: L'ambiente si riferisce all'ambientazione fisica in cui si svolge la scena di ascolto. Questo può essere sia un ambiente interno, come una sala conferenze o un teatro, sia un ambiente esterno, come un parco o un parco divertimenti. Ogni ambiente viene caratterizzato da parametri acustici che descrivono le sue proprietà sonore e influenzano l'intelligibilità del parlato. Tra questi parametri, il tempo di riverberazione (T30) e l'Early Decay Time (EDT) danno un'indicazione del riverbero percepito che ha un impatto sull'intelligibilità del parlato. La Speech Clarity (C50) misura la chiarezza del parlato. Il Direct-to-Reverberant Ratio (DRR) fornisce informazioni sulla distanza percepita della sorgente sonora e su quanto venga percepita come localizzata in una direzione specifica. Infine, lo Speech Transmission Index for Public Address (STIPA) è una misura oggettiva che quantifica l'intelligibilità del parlato in un dato ambiente.
2. *L'ascoltatore*: Rappresenta il punto di vista della scena in cui il soggetto che partecipa al test viene immerso. La posizione e l'orientamento dell'ascoltatore all'interno dell'ambiente sono fondamentali per determinare l'intelligibilità del parlato. Ad esempio, in una scena ambientata in un teatro, l'ascoltatore potrebbe trovarsi al centro della sala, rivolto verso il palco principale.
3. *I parlati target*: Il numero di parlati target viene stabilito in base al tipo di test di intelligibilità che si vuole eseguire. I target possono essere simultanei o alternati, localizzati o diffusi, e possono presentare diverse caratteristiche di direttività. Ogni target è caratterizzato dalla sua posizione e orientamento relativi all'ascoltatore e da parametri acustici misurati nel punto di ascolto, come il livello di pressione sonora equivalente ponderato A (LAeq), lo Speech-Weighted C50 e DRR (calcolati tramite una media pesata in frequenza dei valori di C50 e DRR in base all'importanza delle singole bande di ottava per il parlato) e l'Inter-Aural Cross-Correlation (IACC), che fornisce informazioni sulla spazialità del suono percepito.
4. *Il rumore di mascheramento*: Include suoni che interferiscono con il parlato target, come rumori di fondo, voci di disturbo o suoni ambientali naturali. Più fonti di rumore possono coesistere, influenzando in modo diverso l'intelligibilità del parlato. Si

distinguono quattro categorie principali: (i) *mascheramento energetico (EM)*: si sovrappone fisicamente al parlato in tempo e frequenza, riducendo il rapporto segnale-rumore e rendendo il parlato meno udibile. Tipici esempi sono il rumore del traffico o della ventilazione; (ii) *rumore informativo (IM)*: interferisce a livello cognitivo e si verifica quando da uno a tre discorsi interferenti sono chiaramente intellegibili, condividendo caratteristiche simili al parlato target, rendendo difficile separare le due fonti. Accade, ad esempio, in una conversazione disturbata da altre voci simili; (iii) *multi-talker bubble*: si verifica quando più di tre parlanti interferiscono contemporaneamente, rendendo il rumore interferente non intellegibile, combinando quindi gli effetti di EM e IM; (iv) *speech-shaped noise*: rumore modellato sul parlato che satura completamente il dominio temporale e frequenziale, massimizzando l'effetto di EM. I rumori di mascheramento sono definiti da parametri acustici come LAeq, SW C50, SW DRR e IACC, oltre che dalla loro posizione, direttività e modalità di diffusione.

Nella riproduzione di una scena audiovisiva, ogni elemento segue una struttura gerarchica: l'ambiente determina gli ascoltatori disponibili, la scelta dell'ascoltatore definisce le sorgenti sonore disponibili e le loro posizioni, e così via. Questo approccio modulare garantisce scenari personalizzabili, mantenendo un controllo preciso sulle variabili acustiche.

3. Software per la gestione delle scene

Ogni laboratorio, sia clinico sia di ricerca sull'udito, dispone di strumenti e sistemi specifici, con livelli di complessità differenti a seconda delle esigenze e degli obiettivi dello studio sull'intelligibilità del parlato. Ad esempio, presso il Politecnico di Torino, è stato sviluppato un sistema che utilizza Bidule per la riproduzione di audio spazializzato e Unreal Engine per lo streaming video e quindi dell'immagine visiva correlata, con OSC (Open Sound Control) come protocollo di comunicazione [7]. Grazie alla sua stabilità e affidabilità, OSC è ampiamente utilizzato nella ricerca sull'udito, in particolare nello sviluppo di sistemi per la valutazione e il miglioramento delle protesi acustiche. Inoltre, trova applicazione in sistemi avanzati di rendering acustico, come TASCAR, un ambiente software open-source ampiamente utilizzato nella ricerca sull'udito per la simulazione di scenari uditivi e il processamento di segnali audio spazializzati, sia in tempo reale che offline [6]. Tuttavia, per sfruttare appieno queste potenzialità, specialmente in ambito clinico da personale non esperto nel campo della realtà virtuale, è necessario un software che permetta un'interfaccia intuitiva con tali sistemi.

Il software proposto si pone, quindi, come una possibile soluzione di interfaccia front-end tra l'utente e un sistema di back-end come quello descritto. Inoltre, tramite lo sviluppo di un ulteriore layer intermedio che funga da traduttore, lo stesso software front-end può essere utilizzato come interfaccia utente per un qualsiasi framework back-end alla base, specifico per ogni laboratorio. Si tratta, quindi, di una possibile implementazione della metodologia, e non di un prodotto finito. In altre parole, il software proposto si pone come una possibile soluzione implementativa che consenta la personalizzazione e l'adattamento in base alle esigenze di ogni singolo laboratorio.

Il programma, sviluppato utilizzando il linguaggio Matlab, permette ai clinici di selezionare e personalizzare lo scenario uditivo nel quale si vuole effettuare il test di intelligibilità, tramite un'interfaccia strutturata che guidi il medico passo dopo passo nella procedura. Si ha la possibilità di creare e selezionare profili utente corrispondenti ai pazienti e di navigare tra del-

le schede sequenziali che riflettono la struttura gerarchica descritta dalla metodologia precedentemente definita. Nella prima scheda vengono richiesti i dati del paziente e l'inserimento dell'anamnesi, nella seconda scheda si può scegliere uno tra gli ambienti disponibili, rappresentati da piantine interattive, accompagnate dai modelli 3D e foto dell'ambiente che si vuole riprodurre. Nella terza scheda, vengono visualizzate le possibili posizioni dell'ascoltatore, permettendo di definire il punto di immersione dell'utente nella scena. Per ogni ascoltatore, è possibile configurare l'elevazione e l'azimut, determinando così il punto di vista sonoro all'interno dello scenario. Passando alla finestra successiva, si possono definire il numero, le posizioni e l'orientamento dei target. È quindi possibile inserire uno o più rumori di mascheramento, definendo la loro tipologia; ovvero se localizzate o diffuse, la loro direttività e il loro contenuto. Infine, nell'ultima scheda, è possibile settare i parametri appartenenti al test di intelligibilità specifico selezionato. La selezione della scena avviene in modo progressivo, garantendo che eventuali modifiche a un passaggio non alterino le impostazioni già definite nei passaggi precedenti. Al termine della procedura, un riepilogo finale permette di verificare e confermare tutti i parametri impostati prima dell'esecuzione.

4. Conclusioni

Questo lavoro ha introdotto una metodologia strutturata per la definizione di scenari audiovisivi destinati ai test di intelligibilità del parlato in ambito clinico. La tassonomia sviluppata consente di organizzare in modo sistematico le scene uditive, garantendo coerenza nella selezione e riproduzione delle condizioni d'ascolto. Il software realizzato fornisce ai clinici uno strumento intuitivo per eseguire e gestire i test con precisione, assicurando un'esperienza controllata e adattabile alle diverse esigenze dei pazienti e dei contesti di laboratorio.

Completata la fase di sviluppo della *user mode*, dedicata alla riproduzione delle scene, il lavoro futuro si focalizzerà sulla realizzazione di un modulo software per l'inserimento di nuove scene senza la necessità di modificare direttamente il codice esistente. Questo modulo seguirà lo stesso schema metodologico adottato e guiderà gli utenti nella creazione e integrazione di nuovi scenari, rendendo il processo accessibile anche a chi non ha competenze tecniche avanzate.

5. Bibliografia

- [1] Greenberg, Steven. On the origins of speech intelligibility in the real world. In: *Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels*. 1997. p. 23-32
- [2] Wolters F., Brand T., Kollmeier B., et al., *Common Sound Scenarios: A Context-Driven Categorization of Everyday Sound Environments for Application in Hearing-Device Research*, Journal of the American Academy of Audiology, 27.07 (July 2016), pp. 527-540
- [3] Wu, Yu-Hsiang, et al. Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear and hearing*, 2018, 39.2: 293-30
- [4] Wagener, Kirsten Carola; Hansen, Martin; Ludvigsen, Carl. Recording and classification of the acoustic environment of hearing aid users. *Journal of the American Academy of Audiology*, 2008, 19.04: 348-370
- [5] Macdonald, Justin A., et al. Intelligibility of speech in a virtual 3-D environment. *Human Factors*, 2002, 44.2: 272-286.
- [6] Grimm, Giso. TASCAR: Toolbox for Acoustic Scene Creation and Rendering. Carl von Ossietzky Universität Oldenburg, 2025. Disponibile su: <http://www.tascar.org>. [Accesso: 25 marzo 2025].
- [7] Guastamacchia, A., Rosso, R. G., Puglisi, G. E., Riente, F., Shtrepi, L., & Astolfi, A. (2024, December). Real and Virtual Lecture Rooms: Validation of a Virtual Reality System for the Perceptual Assessment of Room Acoustical Quality. In *Acoustics* (Vol. 6, No. 4, pp. 933-965). Multidisciplinary Digital Publishing Institute.