

VALUTAZIONE DELL'USO DELLA VOCE E DELLA SALUTE VOCALE ATTRAVERSO DISTRIBUZIONI DI PARAMETRI ACUSTICI

TOWARDS VOCAL-BEHAVIOUR AND VOCAL-HEALTH ASSESSMENT USING DISTRIBUTIONS OF ACOUSTIC PARAMETERS

Antonella Castellana *

Dipartimento di Energia, Politecnico di Torino

* Indirizzo dell'autore di riferimento - Corresponding author's address:

Corso Duca degli Abruzzi, 24 – 10129, Torino, Italy

e-mail: antonella.castellana@polito.it; antocast89@gmail.com

(Ricevuto il 19/03/2019, accettato il 17/07/2019)

RIASSUNTO

L'obiettivo di questa tesi è di contribuire all'identificazione di strumenti e requisiti per l'analisi acustica della voce come misura oggettiva di prevenzione e valutazione della disfonia. La prima parte comprende studi sulle distribuzioni di parametri sul carico vocale attraverso monitoraggi degli insegnanti ed esperimenti in laboratorio sugli effetti della riverberazione nell'intensità del parlato. La seconda parte è dedicata alla valutazione della salute vocale: distribuzioni individuali di parametri cespstrali sono stati studiati utilizzando diversi sensori per l'acquisizione del segnale vocale. L'incertezza di ciascun parametro è stata valutata con esperimenti adeguati.

ABSTRACT

The overall aim of this work is to contribute to the identification of tools and requirements related to voice acoustic analysis as objective measure to prevent and assess dysphonia. The first part includes studies on individual distributions of vocal-load parameters, i.e. an in-field study of teachers' voice use and in-lab experiments on the effects of reverberation in speech intensity. The second part is addressed to vocal health assessment: individual distributions of Cepstral Peak Prominence Smoothed from patients and controls were investigated in different speech materials and devices. The uncertainty of each parameter has been evaluated with proper experiments.

Parole chiave: Analisi vocale; parametri acustici; cepstrum.

Keywords: Voice analysis; acoustic parameters; cepstrum.

1. Introduzione

L'OMS definisce limitazioni e le difficoltà che un individuo incontra nell'esecuzione di compiti, e restrizioni tutte le difficoltà che una persona incontra nell'affrontare situazioni di vita quotidiana [1]. La disfonia è un disturbo della comunicazione che impedisce di esprimere il proprio messaggio verbale e le emozioni [2]; in quanto tale, la disfonia può avere un impatto sulla qualità della vita dell'individuo.

Per circa un terzo dei lavoratori mondiali la voce è lo strumento principale durante l'attività lavorativa ed è stato dimostrato che queste persone hanno maggiori probabilità di sviluppare disturbi vocali rispetto ad altri lavoratori [3,4]. Diverse ricerche hanno studiato la frequenza di patologie vocali negli insegnanti [5,6] e tutte affermano che il "carico" vocale e lo "sforzo" vocale sono tra le cause più importanti dei disturbi all'apparato fonatorio.

Di recente sono stati sviluppati dosimetri portatili dotati di microfoni a contatto o accelerometri che rilevano le vibrazioni delle corde vocali per misurare intensità, frequenza fondamentale e tempo di fonazione del parlato. L'importanza di tali dispositivi è legata alla possibilità di raccogliere dati vocali durante le attività lavorative, in quanto il sensore utilizzato per monitorare l'attività vocale ha una limitata sensibilità al rumore di fondo dell'ambiente. Tuttavia, sono stati effettuati pochi studi sul monitoraggio vocale a lungo termine durante l'orario di lavoro ed il più lungo periodo di indagine è stato di due settimane [7]. Pertanto, il primo obiettivo di questo lavoro è quello di:

1. studiare il comportamento vocale degli insegnanti e le relazioni tra l'uso della voce e l'acustica delle classi, attraverso monitoraggi longitudinali della voce durante un intero anno scolastico.

Il paragrafo 2 descrive brevemente la metodologia e include i principali risultati ottenuti come risposte a questo obiettivo di ricerca.

I fattori ambientali appartengono ai fattori di rischio dei problemi vocali professionali. Numerosi studi si sono occupati del diverso comportamento vocale degli oratori in ambienti con diversa acustica, ma questi sono stati spesso incentrati sull'effetto del rumore o della distanza tra oratore e ascoltatore [8]. Pochi dati sono stati pubblicati sul diverso uso della voce in presenza di riverbero e tali ricerche hanno utilizzato solo il microfono in aria per acquisire i segnali vocali [9]. È dunque necessario l'uso di dosimetri portatili in esperimenti di laboratorio per poter valutare i risultati ottenuti in situazioni controllate e comprendere meglio le misure ottenute da monitoraggi sul campo. Inoltre, vi è una mancanza di informazioni sull'incertezza dei parametri vocali stimati con questi dispositivi. Pertanto, il secondo obiettivo di questo lavoro è quello di:

2. analizzare il livello di intensità del parlato in ambienti con un tempo di riverberazione molto basso e molto alto, tenendo conto dell'incertezza dei parametri stimati e usando un microfono a guancia e un analizzatore vocale dotato di microfono a contatto.

Il paragrafo 3 include i principali risultati ottenuti come risposte a questo obiettivo di ricerca.

Il modo in cui si usa la voce ha un ruolo principale nella patogenesi di molti disturbi vocali [10], quindi c'è un crescente interesse da parte dei medici a capire come i pazienti la usino quotidianamente, in modo da identificare i comportamenti non corretti e provare a modificarli. Tale indagine è correlata ai parametri acustici della qualità della voce. Tra i vari parametri acustici proposti per oggettivare il tipo e la gravità della disfonia da segnali acquisiti con microfoni in aria, il Cepstral Peak Prominence Smoothed (CPPS) è stato definito come il più pertinente [11].

Pertanto, il terzo obiettivo di questo lavoro è quello di

3. studiare le distribuzioni CPPS come indicatore di salute vocale sia in vocali sostenute e che in parlato continuo usando microfoni in aria e microfoni a contatto.

Il paragrafo 4 riassume tutte le esplorazioni fatte sulle distribuzioni CPPS: la precisione diagnostica delle statistiche descrittive dalle distribuzioni CPPS ottenute dai diversi materiali vocali e dispositivi, i migliori valori soglia tra i parlanti sani e patologici e le rispettive variabilità, la ripetibilità di tali misure e l'associazione con le valutazioni percettive sono i principali aspetti valutati.

2. Studio longitudinale dell'uso della voce di insegnanti di scuola superiore

2.1 Metodologia

L'attività vocale di 31 e 22 insegnanti di due scuole secondarie è stata monitorata per due settimane lavorative all'inizio (fase 1) e alla fine (fase 2) dello stesso anno scolastico utilizzando un analizzatore vocale basato su sensore a contatto, il Voice Care [12]. I parametri vocali degli insegnanti acquisiti durante le due fasi sono:

- le distribuzioni individuali del livello di voce misurato ad 1 metro di distanza dalle labbra del parlatore, SPL_{1m} e le principali statistiche descrittive;
- le distribuzioni individuali della frequenza fondamentale F_0 e le principali statistiche descrittive;
- il tempo di fonazione, $D_t\%$, ossia la percentuale del tempo totale di monitoraggio trascorso a parlare.

Questi sono stati confrontati e analizzati in relazione ai parametri acustici misurati nelle aule, ossia il tempo di riverberazione (T_{30}) e livello di rumore di fondo (L_{A90}).

La tabella 1 mostra come le aule incluse nella campagna di sperimentazione della scuola A abbiamo una acustica peggiore delle aule considerate per la scuola B.

Tabella 1 – Acustica delle aule: tempo di riverberazione (T_{30}) in condizioni occupate e rumore di fondo (L_{A90}) durante il monitoraggio vocale. L'errore standard è riportato tra parentesi. - Classroom acoustic: reverberation time (T_{30}) in occupied conditions and background noise level (L_{A90}) during voice monitoring. The standard error is reported in brackets.

	<i>School A</i>	<i>School B</i>
T_{30} (s)	1.0 (0.2)	0.5 (0.1)
L_{A90} (dBA)	53.5 (1.0)	50.2 (0.9)

2.2 Principali risultati

La tabella 2 mostra una significativa variazione di alcuni parametri vocali durante le attività didattiche, ma solo nella scuola con i valori di rumore e riverbero più elevati (Scuola A). Gli insegnanti di questa scuola hanno mostrato un aumento del livello di voce medio, $SPL_{mean,1m}$, e una riduzione della percentuale del tempo di fonazione, $D_t\%$, alla fine dell'anno scolastico. Inoltre, nella scuola A la variazione della frequenza fondamentale durante il monitoraggio, F_{0sd} , è diminuita significativamente: questo risultato evidenzia una riduzione della modulazione della voce, che in letteratura è stata spesso correlata a presenza di problemi vocali.

Tabella 2 – Valori medi e errore standard tra parentesi dei parametri vocali per le due fasi dell'anno scolastico (fase 1, fase 2) e per le due scuole. Differenze significative tra le due fasi (p-value <0.05) sono mostrate in grassetto. - Mean values and standard error in brackets of the vocal parameters for the 2 phases of the school year (stage 1, stage 2) and for the 2 schools. Significant differences between the two stages (p-value <0.05) are shown in bold.

	School A			School B		
	Stage 1	Stage 2	Difference	Stage 1	Stage 2	Difference
$SPL_{mean,1m}$ (dB)	68.3 (1.1)	70.6 (1.0)	+2.3 (1.0)*	68.4 (0.8)	68.3 (1.33)	-0.1 (1.5)
$D_t\%$ (%)	50.7 (2.5)	40.4 (2.2)	-10.3 (1.9)*	37.4 (3.1)	38.1 (4.1)	+0.7 (3.4)
$F0_{sd}$ (Hz)	61.7 (3.5)	57.2 (3.5)	-4.5 (1.4)*	57.3 (3.7)	57.3 (2.7)	-1.1 (2.8)

In sintesi, gli insegnanti che lavorano in condizioni acustiche povere usano un livello più alto di voce durante le attività lavorative alla fine dell'anno scolastico e probabilmente riducono il $D_t\%$ al fine di ridurre la sensazione di affaticamento dovuta all'utilizzo di livelli di voce elevati.

Livello di voce e rumore di fondo

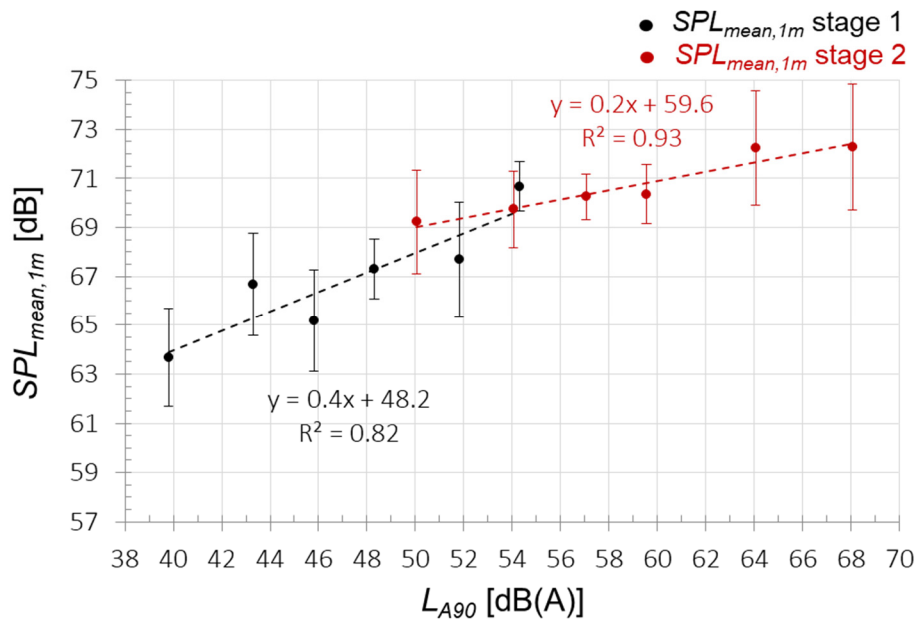


Figura 1 – Regressioni lineari tra il livello di voce ($SPL_{mean,1m}$) e il rumore di fondo (L_{A90}) monitorato durante la fase 1 e la fase 2. Ogni dato sperimentale sul grafico rappresenta il valore medio di 5 coppie - Best fit linear regressions between the voice level ($SPL_{mean,1m}$) and background noise (L_{A90}) monitored during stage 1 and stage 2. Each experimental datum on the graph represents the mean value of an average of 5 pairs.

Per quanto riguarda la relazione tra il livello di rumore di fondo ed il livello di voce, è stato osservato un aumento di $SPL_{mean,1m}$ con l'aumento di L_{A90} . È interessante notare che l'aumento di $SPL_{mean,1m}$ con il rumore all'inizio dell'anno scolastico (0,4 dB/dB) è più

alto di quello osservato alla fine dell'anno (0,2 dB/dB). Solo nel primo caso si può parlare di “effetto Lombard”, come confermato in letteratura da Lane e Tranel [13], Bottalico e Astolfi [14] e Durup et al. [15], i quali hanno trovato relazioni lineari tra SPL e il livello di rumore con pendenze comprese tra 0,3 e 0,7 dB/dB.

La mancanza dell’“effetto Lombard” alla fine dell’anno scolastico può anche indicare che gli insegnanti hanno difficoltà a mantenere lo stesso incremento del loro livello di voce quando il livello di rumore diventa molto alto, cioè superiore a 50 dB (A) come nel presente studio.

Livello di voce e tempo di riverberazione

Le due curve di regressione tra $SPL_{mean,1m}$ e T30 in figura 2 indicano un valore ottimale di riverberazione che supporta la voce delle insegnanti, pari a 0,83 e 0,77 s all’inizio dell’anno scolastico e alla fine, rispettivamente. Questi risultati confermano i risultati di Bottalico e Astolfi [14] e di Puglisi *et al.*[16] ottenuti da monitoraggi in scuole primarie.

La forma delle curve di regressione ottenute è dovuta alla interazione tra l’acustica nelle classi e il livello di voce delle insegnanti: per valori di T30 inferiori di rispetto al valore ottimale la stanza non fornisce un supporto sufficiente alla voce dell’insegnante, quindi questa tende ad utilizzare un livello di voce più elevato. Per valori di T30 superiori rispetto al valore ottimale, invece, è possibile che un’elevata riverberazione produca un livello più elevato di rumore nella stanza, che a sua volta porta gli insegnanti ad alzare il livello di voce, confermando la tendenza all’“effetto Lombard”.

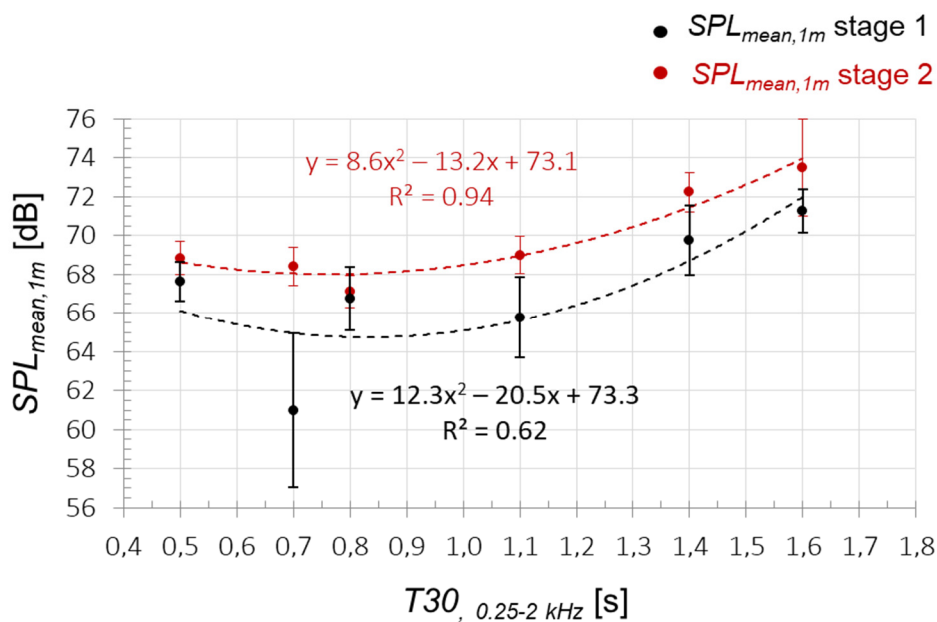


Figura 2 – Regressioni lineari tra il livello di voce ($SPL_{mean,1m}$) e il tempo di riverberazione (T30) durante la fase 1 e la fase 2. Ogni dato sperimentale sul grafico rappresenta il valore medio di 5 coppie - Best fit linear regressions between the voice level ($SPL_{mean,1m}$) and reverberation time (T30) of stage 1 and stage 2. Each experimental datum on the graph represents the mean value of an average of 5 pairs.

3. Studio dell'intensità del parlato in ambienti con un tempo di riverberazione molto basso e molto alto

3.1 Metodologia

Gli esperimenti sono stati effettuati nelle camere semi-anecoica e riverberante dell'Istituto Nazionale di Ricerca Metrologica (I.N.Ri.M.) di Torino, caratterizzate da un tempo di riverberazione alle medie frequenze di 0,11 (s.d. 0,01) e 7,38 (s.d. 1,61), rispettivamente.

I compiti di ciascun parlatore sono stati raccontare una storia breve della quale si aveva conoscenza e descrivere una mappa, ad un ascoltatore seduto di fronte a lui/lei nella camera semi-anecoica e poi nella camera riverberante. La variazione dei parametri vocali è stata rilevata attraverso l'uso del microfono omnidirezionale ad archetto Mipro MU-55HN (Chiayi, Taiwan) e di Voice Care. Il campione di parlatori era composto da 44 soggetti, 21 maschi e 23 femmine. La tabella 3 riporta il numero di soggetti che hanno intrapreso i vari esperimenti in entrambi le camere.

Tabella 3 – Numero totale di soggetti, femmine (F) e maschi (M), che hanno partecipato agli esperimenti per tipo di microfono indossato e per compito eseguito – Total number of participants to the experiment, females (F) and males (M), for each device and speech task.

	<i>Contact device</i>			<i>Headworn microphone</i>		
	<i>F</i>	<i>M</i>	<i>Overall</i>	<i>F</i>	<i>M</i>	<i>Overall</i>
Free speech	8	15	23	16	13	29
Map	5	10	15	14	14	28
<i>Overall</i>	13	25	38	30	27	57

Prima di iniziare a parlare, a ciascun soggetto è stato chiesto di ripetere la vocale /a/ e di colpire allo stesso tempo l'ECM del dispositivo Voice Care, in modo che il segnale acquisito dal microfono contatto e il segnale acquisito registrato dal microfono ad archetto Mipro MU-55HN si potessero sincronizzare.

Per ogni campione di parlato sono stati ottenuti il livello di pressione sonora medio, SPL_m , la moda, SPL_{mode} , e il livello di pressione sonora equivalente, SPL_{eq} . Nel caso del dispositivo Voice Care, il livello di pressione sonora equivalente stimato dalla vibrazione della pelle, è stato determinato in accordo con Švec et al. [17]. Nel caso del microfono ad archetto Mipro MU-55HN, è stato determinato il livello di potenza sonora (PWL) in base a una procedura descritta in breve di seguito. È stato prima determinato un fattore di correzione, G_{refl} , dovuto all'aumento del livello di pressione sonora equivalente complessiva al microfono ad archetto nella camera riverberante rispetto alla camera semi-anecoica. In modo analogo a quanto descritto da Brunskog et al. [17] è stata determinata G_{DIST} , cioè la differenza tra il livello di pressione sonora equivalente corretto dalle riflessioni e il livello di potenza sonora del simulatore, eseguendo misure in camera riverberante in accordo con la norma ISO 3743- 2 [19]. Il livello di potenza sonora di ogni soggetto in camera riverberante è stato infine stimato come differenza tra il livello equivalente corretto dalle riflessioni nella camera riverberante e G_{DIST} ; mentre il livello di potenza sonora di ogni soggetto in camera semi-anecoica è stato stimato come la differenza tra il livello di pressione sonora equivalente nella camera semi-anecoica e G_{DIST} .

3.2 Principali risultati

Per quanto riguarda i monitoraggi effettuati con il Voice Care, che utilizza una finestra di analisi di 30 ms ed elimina le pause di silenzio nel parlato, nella camera semi-anechoica (sa) si è osservata la tendenza ad avere livelli di pressione sonora superiori a quelli registrati in camera riverberante (r). Come mostra la tabella 4, valori medi superiori di 2,0 dB, 1,9 dB e 2,4 dB sono stati registrati nella camera semi-anechoica rispetto alla camera riverberante nel caso di descrizione di una mappa. Queste differenze sono significative in quanto i loro valori sono superiori alla rispettiva incertezza estesa per le differenze tra i livelli di intensità del parlato, che è pari a 1,4 dB per $\Delta\text{SPL}_{\text{eq}}$ e ΔSPL_m , e a 1,7 dB per $\Delta\text{SPL}_{\text{mode}}$. Tali valori sono stati stimati come ulteriore lavoro della presente tesi: ciascun contributo di incertezza della misura è stato valutato con un opportuno esperimento e combinato a tutti gli altri seguendo la Guida all'espressione dell'incertezza di misura [20].

Nel caso del compito con un maggiore intento comunicativo, la descrizione della mappa, si è evidenziato quindi un maggiore sforzo vocale nella stanza povera di riflessioni. Lo stesso comportamento non è risultato significativo nel caso di parlato libero.

Nel caso del microfono a guancia, per il quale è stato utilizzato un intervallo di registrazione di 1 s, non sono state riscontrate differenze significative né nei livelli di pressione del suono della voce né nei livelli di potenza sonora. Sono state condotte indagini per determinare come un intervallo di registrazione di 30 ms, la cui lunghezza è paragonabile alla pausa inter-sillabica, possa influenzare la stima dei parametri SPL durante l'utilizzo del microfono a guancia. È consigliato un intervallo di registrazione di 1 secondo per la stima dei parametri SPL con microfoni in aria, in quanto li rende meno interessati dal rumore registrato nelle pause del parlato. In particolare, il parametro meno influenzato è SPL_{eq} . Questo studio ha dimostrato che i dispositivi basati su microfoni a contatto sono più appropriati dei microfoni in aria per il monitoraggio vocale in campo, poiché sono in grado di rilevare solo l'attività delle corde vocali, senza alcun artefatto esterno.

Tabella 4 – Valori medi (in alto) e errore standard (in basso) di SPL_{eq} , SPL_m , e SPL_{mode} , stimati con il microfono a contatto nella camera anecoica (sa) e riverberante (r) e differenza dei livelli tra le due camere ($\Delta_{\text{sa-r}}$) - Average value (upper cells) and standard error (lower cells) of SPL_{eq} , SPL_m , and SPL_{mode} , estimated with the contact device in the semi-anechoic (sa) and reverberant (r) rooms, and level differences between the two rooms ($\Delta_{\text{sa-r}}$).

Task	Subj.	SPL_{eq}		SPL_m		SPL_{mode}		PWL		$\Delta\text{SPL}_{\text{eq}}$	ΔSPL_m	$\Delta\text{SPL}_{\text{mode}}$	ΔPWL
		sa	r	sa	r	sa	r	sa	r				
Free speech	29	94.3	93.5	92	91.7	94.1	93.9	71	69.8	0.8	0.3	0.2	1.2
		0.8	0.9	0.8	0.9	0.9	0.8	0.8	0.9				
Map	28	94.7	94.3	89.8	90.9	94.8	95.2	71.3	70.7	0.4	-1.1	-0.4	0.6
		0.9	0.9	0.9	0.8	1	0.9	0.9	0.9				

Tabella 5 – Valori medi (in alto) e errore standard (in basso) di SPL_{eq}, SPL_m, SPL_{mode}, e PWL stimati con il microfono a contatto nella camera anecoica (sa) e riverberante (r) e differenza dei livelli tra le due camere (Δ sa-r) - Average value (upper cells) and standard error (lower cells) of SPL_{eq}, SPL_m, SPL_{mode} and PWL estimated with the contact device in the semi-anechoic (sa) and reverberant (r) rooms, and level differences between the two rooms (Δ sa-r).

Task	Subj.	SPL _{eq}		SPL _m		SPL _{mode}		PWL		Δ SPL _{eq}	Δ SPL _m	Δ SPL _{mode}	Δ PWL
		sa	r	sa	r	sa	r	sa	r				
Free speech	29	94.3	93.5	92	91.7	94.1	93.9	71	69.8	0.8	0.3	0.2	1.2
		0.8	0.9	0.8	0.9	0.9	0.8	0.8	0.9				
Map	28	94.7	94.3	89.8	90.9	94.8	95.2	71.3	70.7	0.4	-1.1	-0.4	0.6
		0.9	0.9	0.9	0.8	1	0.9	0.9	0.9				

La tabella 5 mostra come nel caso del microfono a contatto tutte le differenze sono inferiori alla rispettiva incertezza estesa per le differenze tra i livelli, pari a 1,9 dB per Δ SPL_{eq}, 1,8 dB per Δ SPL_m e pari a 2,0 dB per Δ SPL_{mode}. Cautelativamente, nel caso di Δ SWL l'incertezza potrebbe essere considerata la stessa di Δ SPL_{eq}, da cui si ottiene Δ SWL, in cui i contributi di Grefl e Gdist non sono considerati.

4. Distribuzioni CPPS come indicatore di salute vocale

Il Cepstral Peak Prominance Smoothed, CPPS, definito in [21], misura la differenza in decibel tra il picco del cepstrum e del corrispondente valore sulla retta di regressione, calcolata sul grafico medio ottenuto da cepstra di più *frames* del campione vocale. Il CPPS è quindi una misura del grado di armonia dello spettro del campione vocale: più periodico è il segnale vocale, maggiore è la regolarità delle armoniche dello spettro e maggiore risulta il suo valore. Questo paragrafo riassume le esplorazioni fatte sulle distribuzioni di CPPS come descrittori dello stato di salute vocale in diversi materiali vocali e utilizzando diversi tipi di microfono per il monitoraggio della voce.

4.1 Metodologia

Il materiale vocale analizzato in questo studio è stato ottenuto da prove sperimentali eseguite nel reparto di Otorinolaringoiatria dell'Ospedale Universitario Molinette San Giovanni Battista di Torino, in un ambiente con un limitato rumore di fondo. In particolare, indossando sia il microfono a guancia che il microfono a contatto (ECM):

1) a ciascuno dei 41 pazienti e 35 soggetti di controllo è stato chiesto di pronunciare la vocale /a/ ad un livello di comfort, fino a quando non avessero avvertito la necessità di effettuare nuovamente un respiro;

2) a ciascuno dei 72 pazienti e 39 soggetti di controllo è stato chiesto di leggere a voce alta un brano standardizzato e di parlare liberamente di un argomento a loro noto per un minuto.

Due otorinolaringoiatri hanno eseguito l'esame di laringostroboscopia a tutti i soggetti. Le patologie vocali de gruppo dei pazienti sono principalmente edema, cisti, laringite cronica, ipostenia vocale, disfonia postoperatoria, nodulo, disturbo neurologico, disfonia funzionale.

Per valutare quanto le occorrenze del CPPS siano in grado di discriminare una voce disfonica da una sana, le statistiche descrittive delle distribuzioni sono state confrontate con l'esito della laringostroboscopia attraverso modelli di regressione logistica multipla.

4.2 Principali risultati

Per quanto riguarda le vocali acquisite con il microfono in aria, il quinto percentile (CPPS_{5prc}) è risultato la migliore statistica descrittiva per le distribuzioni CPPS in grado di discriminare voci sane e patologiche. Il rispettivo modello logistico ha mostrato un forte potere discriminante (Area Under the Curve, AUC = 0.95) e una soglia di discriminazione di CPPS_{5prc} = 15.0 dB (intervallo di confidenza al 95%, IC, di 0.7 dB), con valori più bassi che indicano lo stato di voce non salutare (Fig. 3).

Per quanto riguarda le vocali sostenute acquisite con l'ECM, invece, la deviazione standard (CPPS_{std}) è il parametro migliore che separa i due gruppi. Il rispettivo modello logico empirico ha un buon potere discriminatorio, con AUC di 0,87 e una soglia di discriminazione di CPPS_{std} = 1,1 dB (IC di 0,2 dB), con valori maggiori per le voci patologiche. A differenza dei risultati di Mehta *et al.* [22], il metodo proposto è in grado di discriminare la voce sana e patologica sia dal microfono in aria sia da un microfono di contatto. La variabilità intra-soggetto dei due parametri di CPPS è risultata maggiore nel gruppo dei pazienti rispetto a quella di controllo: i rispettivi valori sono pari a 0,8 dB e 0,5 dB per CPPS_{5prc} e 0,3 dB e 0,2 dB per CPPS_{std}. Questo risultato evidenzia la limitata capacità dei pazienti con patologia vocale nella produzione vocale.

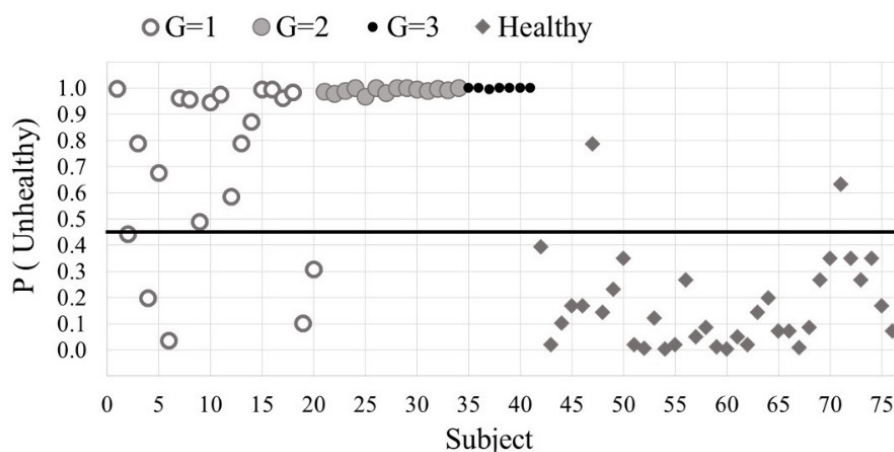


Figura 3 – Valori del miglior modello di regressione logistica, in termini di probabilità di avere una voce patologica, per vocali acquisite con il microfono a guancia. I cerchi indicano il gruppo di pazienti (con diversi gradi G di disfonia); i diamanti rappresentano i soggetti sani. La linea indica il valore di soglia (0,44, cioè 15 dB) - Fitted values of the best logistic regression model, in terms of probability of having unhealthy voice, for vocalizations acquired with the headworn microphone. Circles indicate the patient group (with different overall grade G of dysphonia); diamonds represent the controls. The bold line indicates the threshold value (0.44, i.e. 15 dB).

Con l'obiettivo di fornire linee guida che rendano affidabili le stime dei parametri CPPS, è stata eseguita un'analisi delle principali quantità di influenza del parametro. I risultati ottenuti hanno evidenziato che la frequenza fondamentale e il rapporto segnale-

rumore (SNR) dei segnali acquisiti possono significativamente influenzare la discriminazione tra voci sane e patologiche. Per questo motivo, è importante limitare il campo di utilizzo della frequenza fondamentale, ad esempio fornire un tono di riferimento al soggetto prima di eseguire il compito richiesto e evitare una grande differenza in SNR durante la campagna sperimentale. Ulteriori indagini sono state fatte al fine di stimare l'effetto del contenuto di frequenza dello spettro del segnale sui parametri CPPS. Come risultato di questa analisi, si può affermare che si ottiene una stima affidabile dei parametri $CPPS_{5prc}$ e $CPPS_{std}$ purché il contenuto in frequenza dello spettro non sia inferiore a 5 kHz. Ciò giustifica la potenza di discriminazione inferiore ottenuta per il microfono a contatto che mostra un contenuto di frequenza di circa 3,5 kHz.

Per quanto riguarda il parlato continuo acquisito con il microfono a gancia, il $CPPS_{95prc}$, è risultato il miglior parametro nella discriminazione tra voci sane e patologiche, sia per la lettura che per il discorso libero (AUC di 0.86). La soglia di discriminazione è pari a 18,0 dB (IC di 0,6 dB), dove valori più bassi indicano un'alta probabilità di avere una voce patologica. Per quanto riguarda i materiali vocali acquisiti con l'ECM, non è stato ottenuto un potere di discriminazione ragionevole (AUC superiore a 0,80).

Per entrambi i tipi di microfono utilizzati è stata trovata una peculiare caratteristica nelle distribuzioni di CPPS: la tendenza centrale e la forma delle distribuzioni individuali sono le stesse sia per la lettura che per il parlato libero. Tale caratteristica intrinseca è stata nominata “impronta vocale” del CPPS (Fig.4).

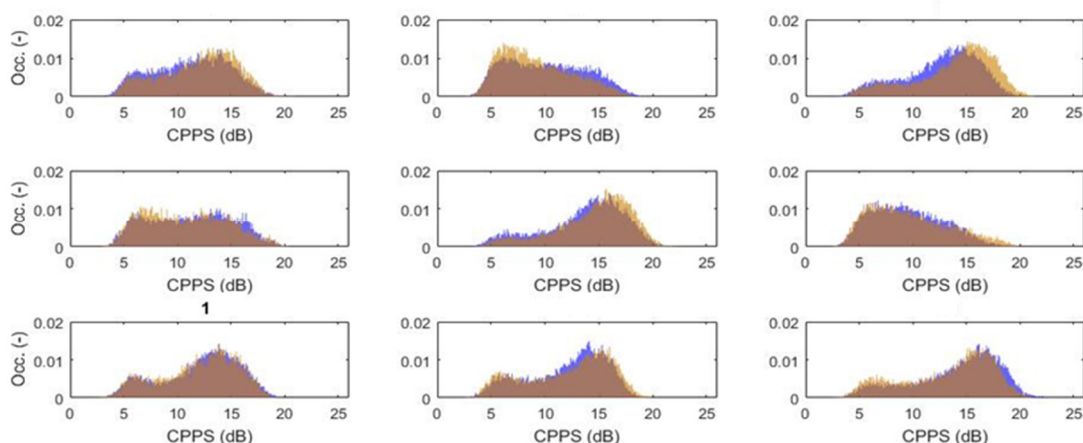


Figura 4 – Distribuzioni di CPPS ottenute dall'analisi della lettura (blu) e del discorso libero (marrone) acquisite da 9 pazienti con il microfono a guancia - CPPS distributions obtained from the analyses of reading (blue) and free speech (brown) tasks acquired with the headworn microphone from 9 patients.

Conclusioni

L'intera ricerca fatta per questa tesi di dottorato pone le basi ad un approccio olistico per la valutazione della salute vocale, in cui i mezzi sono sensori a contatto che possono essere utilizzati durante le attività quotidiane e gli output sono distribuzioni di parametri acustici che sono in grado di descrivere lo stato della voce in modo efficiente. Questo approccio soddisfa sia le esigenze dei medici che dei pazienti, in quanto può fornire un'enorme quantità di dati ed un feedback continuo. In ricerche future occorre utilizzare

campioni di dati più numerosi e omogenei per studiare le distribuzioni dei parametri utilizzando metodi di clustering.

A conclusione del percorso di dottorato, PR.O.VOICE S.r.l. (Torino, Italia) ha sviluppato strumenti specifici in grado di accogliere tutti gli output della tesi, ovvero un'applicazione per smartphone combinata con un microfono a contatto inserito in un collare (Vocal Holter App) e la versione professionale, un dispositivo indossabile (il Volcal Holter Med). Sono in corso monitoraggi a lungo termine della voce di insegnanti utilizzando l'app e sarà possibile conoscere sia i parametri relativi al carico vocale, come le distribuzioni SPL e F0, sia il loro stato di salute vocale utilizzando le distribuzioni CPPS, durante le ore lavorative.

Conclusions

The whole research done for this Ph.D. thesis puts the basis for a holistic approach of vocal health assessment, where the tools are contact sensors that can be used during daily activities and the outputs are distributions of parameters that are able to describe the voice status efficiently. This approach meets both clinicians' and patients' needs as it can provide a huge amount of data and continuous feedback. Future researches need to use larger and more homogeneous data samples to investigate on distributions using clustering methods.

Specific tools able to merge all the outputs of this Ph.D. thesis have been developed by PR.O.VOICE S.r.l. (Turin, Italy), that are, a smartphone application combined with a cheap contact microphone embedded in a collar (the Vocal Holter App), and the professional version, a wearable device (the Volcal Holter Med). In-field long-term monitorings of teacher's voice using the app are now in progress and it will be possible to know both vocal load related parameters, such as SPL and F0 distributions, and their vocal health status using CPPS distributions, during the working hours.

Bibliografia

- [1] World Health Organization (2001). International Classification of Functioning, Disability and Health: ICF.
- [2] Vilkmann E. (2004). Occupational safety and health aspects of voice and speech Professions. *Folia Phoniatrica et Logopaedica*, 56(4), pp.220–253.
- [3] Behlau M., Zambon F., Guerrieri A.C., Roy N. (2012). Epidemiology of voice disorders in teachers and nonteachers in Brazil: prevalence and adverse effects. *Journal of voice*, 26(5), pp.665–e9.
- [4] Godall P., Gassull C., Godoy A., Amador M. (2015). Epidemiological voice health map of the teaching population of granollers (Barcelona) developed from the EVES questionnaire and the VHI. *Logopedics Phoniatrics Vocology*, 40(4), pp.171-178.
- [5] Simberg S., Sala E., Vehmas K., Laine A. (2005). Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *Journal of Voice*, 19(1), pp.95–102.
- [6] Cantor Cutiva L.C., Vogel I., Burdorf A. (2013). Voice disorders in teachers and their associations with work-related factors: a systematic review. *Journal of Communication Disorders*, 46(2), pp.143–155.
- [7] Hunter E.J, Titze I.R. (2010). Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *Journal of Speech, Language, and Hearing Research*, 53(4), pp.862–875.

- [8] Cheyne H.A, Kalgaonkar K., Clements M., Zurek P. (2009). Talker-to-listener distance effects on speech production and perception. *The Journal of the Acoustical Society of America*, 126(4), pp.2052–2060.
- [9] Pelegrín-García D., Smits B., Brunskog J., Jeong C. (2011). Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4), pp.1981–1990.
- [10] Hillman R.E., Holmberg E.B., Perkell J.S, Walsh M., Vaughan C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech, Language, and Hearing Research*, 32(2), pp.373–392.
- [11] Maryn Y., Roy N., De Bodt M., Van Cauwenberge P., Corthals P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America*, 126(5), pp.2619–2634.
- [12] Carullo A., Vallan A., Astolfi A. (2013). Design Issues for a Portable Vocal Analyzer. *IEEE Trans. on instrumentation and measurement* 62(5), pp.1084-1093.
- [13] Lane H., Tranel B. (1971). The lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, 14(4), pp.677–709.
- [14] Bottalico P., Astolfi A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *The Journal of the Acoustical Society of America*, 131(4), pp.2817–2827.
- [15] Durup N., Shield B., Dance S., Sullivan R. (2015). An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Building Acoustics*, 22(3-4), pp. 225–241.
- [16] Puglisi G.E., Astolfi A., Cantor Cutiva L.C., Carullo A. (2017). Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America*, 141(1), pp. 441–452.
- [17] Švec J.G, Titze I.R., Popolo P.S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America*, 117(3), pp.1386–1394.
- [18] Brunskog J., Christian Gade A., Payá Bellester G., Calbo L.R. (2009). Increase in voice level and speaker comfort in lecture rooms. *The Journal of the Acoustical Society of America*, 125(4), pp.2072–2082.
- [19] ISO (2010). ISO 3741:2010, Acoustics - Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for reverberation test rooms.
- [20] CEI (2016). CEI UNI 70098-3:2016 – Incertezza di misura – Parte 3: Guida all'espressione dell'incertezza di misura, Comitato Elettrotecnico Italiano, Milano.
- [21] Hillenbrand J., Houde R.A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), pp. 311–321.
- [22] Mehta D.D, Van Stan J.H., Hillman R.E. (2016). Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(4), pp.659–668.