

08 Beyond the mask

Quest'anno, a seguito di una situazione di emergenza sanitaria, le nostre abitudini quotidiane sono state stravolte.

I nuovi dispositivi di sicurezza, che coprono il volto, come le mascherine, si sono inseriti come un oggetto estraneo. Ciò ha rappresentato un problema enorme per le persone non udenti, perché limita moltissimo la possibilità di comunicare. I sordi hanno bisogno di leggere il labiale, per poter decifrare le parole che non sentono. L'intento è rendere visibile il linguaggio verbale e non verbale per includere le persone con difficoltà uditive.

Ilenia Balella



#mask
#AR
#p5speech
#parole
#silenzio

github.com/dsii-2020-unirsm
github.com/ileniab

a destra
immagine evocativa,
unsplash.com



Ricerca.

La mascherina è entrata nella nostra quotidianità, inserendosi come un oggetto estraneo che filtra e nasconde parte della nostra natura espressiva e comunicativa.

A seguito di questo periodo assolutamente particolare, come stanno reagendo le comunità minoritarie? Sono già partite delle iniziative che promuovono la realizzazione e produzione di mascherine trasparenti, che consentono alle persone non udenti di poter vedere il movimento delle labbra. *DHH Project, Deaf and hard of hearing project*, è un progetto partito negli Stati Uniti, che si è trasformato in pochi giorni in una raccolta fondi per finanziare la produzione di queste particolari mascherine. Negli USA non è la prima volta che se ne parla, già nel 2017 era entrata in commercio una mascherina chirurgica trasparente, *The Communicator Clear Window Mask* prodotta da *Safe 'n Clear*.

All'interno della comunità sorda ci sono diverse sfumature, ed è bene sottolinearne qualcuna. Non tutti conoscono la lingua dei segni, magari perché sono diventati sordi ad una certa età, oppure hanno solo un livello lieve di sordità ed imparano spesso a leggere solamente il labiale, (definiti "sordi oralisti"). Invece il "sordo segnante" usa la lingua dei segni per comunicare. In Italia la perdita di udito è suddivisa in quattro fasce: lieve, media, grave, profonda.^[1] Nelle prime due fasce la legislazione italiana riconosce un grado più o meno elevato di invalidità, e soltanto da una perdita superiore di 75 decibel si riconosce lo statuto di "sordo", a coloro che hanno perso l'udito prima dei 12 anni, e per un'ipoacusia pari o superiore a 60 dB all'orecchio migliore (cfr. legge 381/70 pag. 27).

Oliver Sacks nel libro "Vedere voci" parla delle differenze tra i sordi postlinguistici, ovvero coloro che hanno perso l'udito successivamente all'acquisizione del linguaggio vocale, e i prelinguistici, che sono nati sordi e non hanno avuto la possibilità di udire alcun suono. La differenza principale tra i due è che

[1] Decreto Ministeriale 5.2.1992

in alto
mascherina DHH project, 2020

in basso
The Communicator Clear
Window Mask, 2017

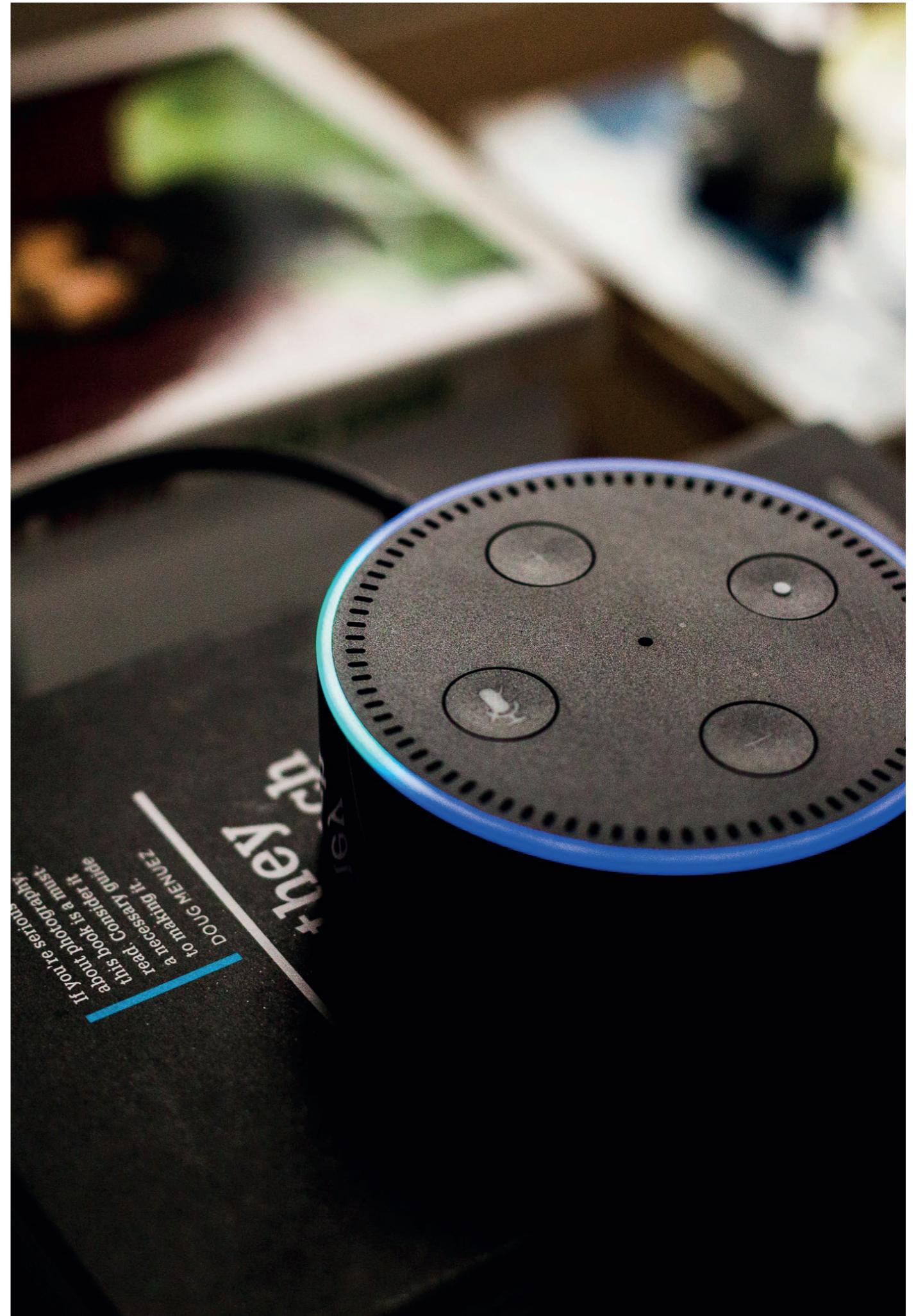


i primi, nella loro memoria hanno ancora il ricordo del rumore, e grazie ad un fenomeno chiamato “voce fantasma”, la loro mente proietta automaticamente dei suoni, che hanno in memoria, sull’esperienza visiva. Ovvero riescono a sentire per esempio il suono delle foglie mosse dal vento se guardano un albero, oppure il suono di una voce familiare. Ciò non accade per i sordomuti, che vivono in un mondo di totale assenza di suono. Solo in Italia ci sono circa 5 milioni di audiolesi, di cui 70.000 sordi prelinguali.

Senza la possibilità di vedere il labiale, non hanno strumenti per leggere. Perché avremmo bisogno di una app che trascriva il testo in realtà aumentata? Perché non utilizzare già le funzionalità che offre il sistema di Google *Speech-to-speech translation systems*, che rileva in automatico il parlato e lo traduce nella lingua di preferenza? Purtroppo il traduttore di testo di Google è stato criticato per aver fatto spesso traduzioni incomplete o errate, ed è spesso descritto come approssimativo e talvolta inutile. L’incapacità di Google Translate di produrre una buona grammatica è un problema informatico, ed un limite tecnologico ancora in via di sviluppo. La lingua parlata ha infinite sfumature, ed è estremamente difficile tenerne conto nella programmazione di un software. I modelli per il *machine learning* dovranno essere perciò sempre più accurati.

Quando si parla di *Speech technology* si intende qualsiasi tecnologia che abbia a che fare con la voce. Ci sono diverse categorie tra cui *speech recognition*, *speech verification*, *real time speech to text conversion*, *interactive voice response* (IVR), *speech synthesis* e *speech analytics*. La tecnologia TTS (*text-to-speech*) converte il testo scritto in una sintesi vocale. È spesso utilizzata nell’ambito delle disabilità visive, da persone con dislessia o altre difficoltà di lettura. Inoltre la sintesi vocale viene impiegata anche nell’intrattenimento come giochi e animazioni. La NLP, *Natural Language Processing*, è la capacità di un elaboratore elettronico di comprendere un linguaggio nella sua complessità. Infatti adotta un

a destra
immagine evocativa,
unspalsh.com



metodo di *Deep Learning* per l'analisi semantica e sintassi. Riconosce il contesto d'uso di una parola, e utilizza un database per determinare la semantica.

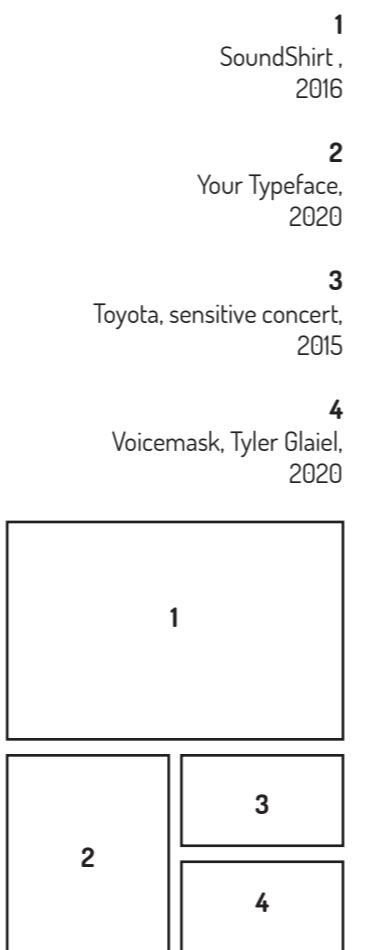
Casi studio.

SoundShirt è un progetto realizzato da una società di moda, *CuteCircuit*, con sede a Londra. Si tratta di una maglia che sfrutta la sensazione aptica, per far esperire la musica con altri sensi. È stata realizzata associando determinati strumenti alle diverse parti del corpo, in modo tale da avere delle vibrazioni distinte a seconda del tipo di suono. Per le persone sordi, e non solo, è un nuovo modo di percepire la musica e di sperimentare fisicamente nuove sensazioni anche nel mondo virtuale.

Your Typeface, è un progetto che esplora il mondo dei font variabili. Per questo progetto è stato creato un sito, che permette all'utente di interagire con la variabilità del font utilizzando la webcam. Il font reagisce alle espressioni del viso in tempo reale, l'utente giocherà contraendo tutti i muscoli facciali per scoprire il funzionamento e cercare il carattere tipografico preferito da poter scaricare. Il font varia in base a diversi parametri: l'apertura in altezza della bocca regola il peso del font, la larghezza invece regola la compattezza.

Toyota sensitive concert, realizzato in collaborazione con l'Ente Nazionale Sordi e la casa automobilistica, è stato un evento multisensoriale realizzato per avvicinare il mondo musicale alla comunità sorda attraverso la sensazione tattile. Questo evento ha avuto luogo nel 2015 al Torino Jazz Festival e la Toyota ha realizzato delle sfere che connettono la musica a stimoli tattili e visivi. Le sfere erano collegate agli strumenti dei musicisti che suonavano sul palco. Si potevano riprodurre vibrazioni relative alle diverse frequenze degli strumenti musicali e le sfere si illuminavano a tempo di musica.

Voicemask è un progetto realizzato da Tyler Glaiel, un programmatore e game designer statunitense.



Your Typeface

A digital experience by Overtone and Set Snail



Ha realizzato una maschera con una matrice led 8x8, comandata da Arduino, che imita il movimento della bocca. In base a determinate variabili, ha assegnato un “disegno led” ad un range di volume della voce, in modo tale da avere un movimento fluido dei led. C’è anche la possibilità di far “sorridere” la maschera emettendo il suono “pop”.

Progetto | cos’è

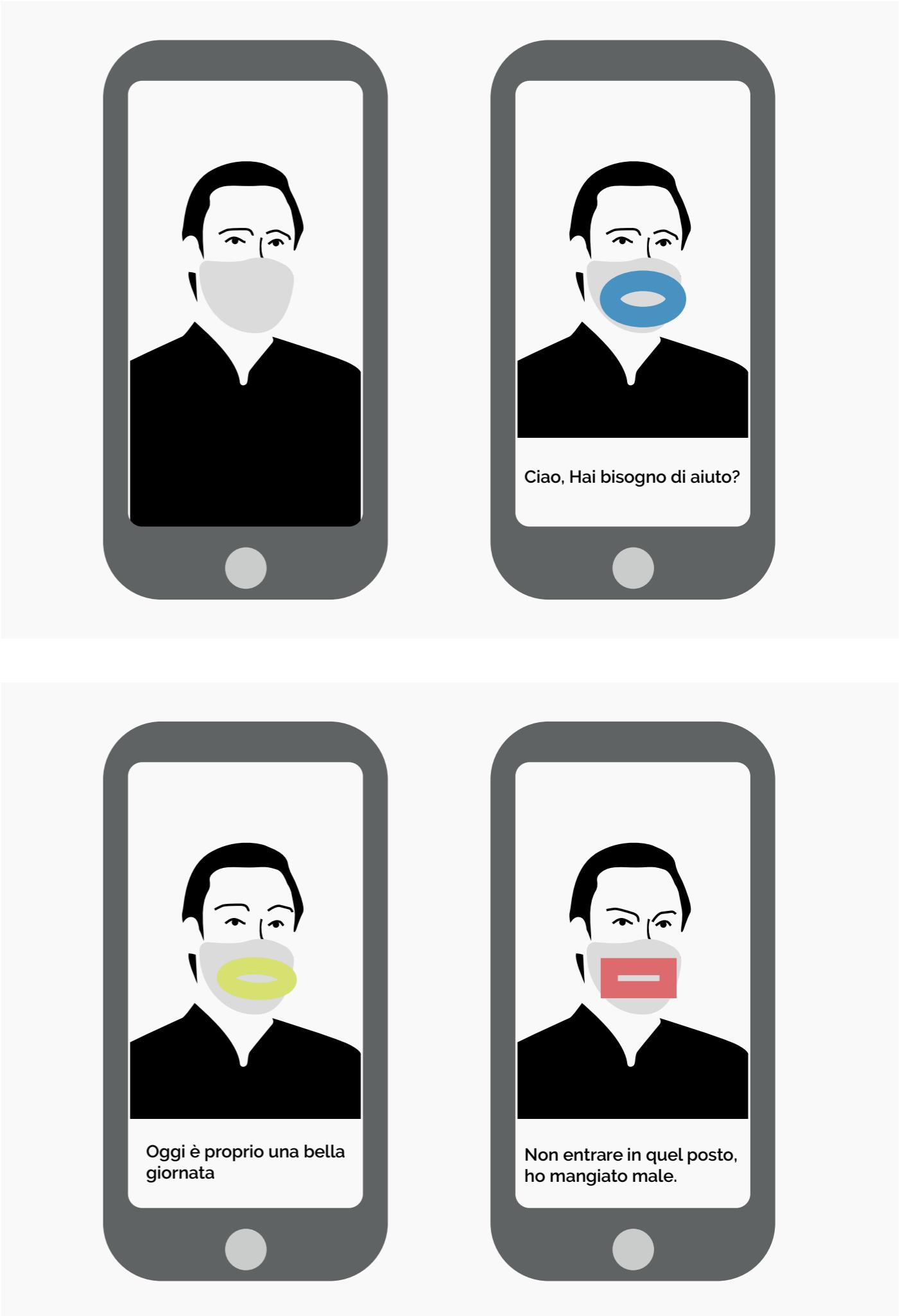
Avendo come target ampio range di persone con difficoltà uditive, ho optato per una app scaricabile ed utilizzabile da smartphone o tablet, che sfrutta la telecamera e il microfono del dispositivo, per raffigurare in diretta in realtà aumentata il linguaggio verbale, per mezzo di una sottotitolazione, e la comunicazione non verbale che viene espressa tramite forme e colori che si animano sul volto dell’interlocutore.

L’esigenza progettuale di utilizzare una app nasce dalla necessità e dalla possibilità di usufruirne più facilmente, ed essere accessibile in qualsiasi momento della giornata, per strada, in ufficio, in casa, in un supermercato, al cinema. Attraverso l’utilizzo della sottotitolazione, l’utente potrà comprendere il discorso, anche se l’interlocutore indossa una mascherina, e riuscire ad interagire in sicurezza. L’aggiunta di giochi di forme e colori, non solo favorisce la trasmissione di valori semantici ed emozionali all’interno della comunicazione, ma restituisce umanità al volto. L’idea è quella di raffigurare una forma ovale che simula il movimento di una bocca che parla. Infine inquadrando il volto della persona che parla, c’è la possibilità di mantenere un contatto visivo con l’interlocutore.

Progetto | come funziona

Le parole non bastano a trasmettere globalmente quel che si vuole comunicare agli altri. Degli studi hanno dimostrato che la comunicazione non verbale ha un’incidenza del 55% sulla trasmissione del messaggio (Albert Mehrabian, 1972).

in alto
immagine esplicativa
del progetto
in basso
immagine esplicativa
del progetto



Perciò per veicolare le informazioni non verbali celate dalla mascherina, si utilizza il colore e la forma di un oggetto, che riesce ad aggiungere pathos generando empatia, scavalcando il muro inespressivo generato dalla mascherina.

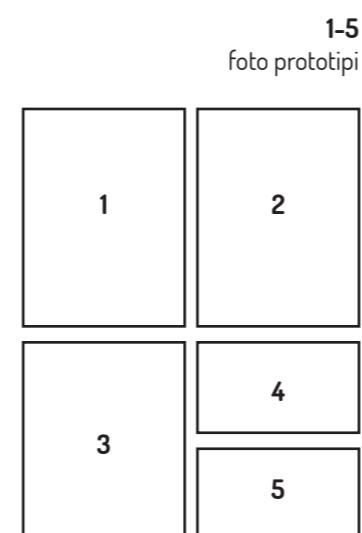
Utilizzando la telecamera di un dispositivo mobile si inquadra il volto dell'oratore l'algoritmo di *FaceApi* permette di inquadrare e riconoscere il volto, la libreria di *P5 speech* permette di registrare e trascrivere il testo. Mentre l'interlocutore parla, si attiva la *Sentiment Analysis*, che cambia la forma al riquadro sul volto, in base alla semantica del testo, rispettivamente al range positivo e negativo. Per raffigurare il movimento di una bocca che parla assegnerò delle variabili che in base alla presenza di determinate vocali il riquadro si allargherà o stringerà.

A determinare il cambiamento del colore (arbitrario), è l'espressione facciale, l'emozione verrà rilevata dalla versione completa di *FaceApi*, allenata sul volto coperto dalla mascherina. Infine il tono e l'intensità della voce potrà variare la dimensione del font.

Esiste un progetto chiamato *Empath Web API*, che utilizza il linguaggio python e riesce a determinare l'emozione in base al tono e alla frequenza della voce, come rabbia, gioia e tristezza.

Progetto | prototipi

Il prototipo è composto da quattro moduli autonomi tra loro collegati. La libreria di *p5 speech* (ascolta e registra una conversazione), l'algoritmo di *Sentiment Analysis* (analizza il testo e fa una predizione che va da 0 a 100, del quale ho determinato 3 parametri di valore positivo, neutro e negativo), *FaceApi* (che sfrutta le coordinate del volto per calcolare dove disegnare un riquadro) ed *Emotion recognition*, oltre ad un modello allenato a riconoscere una mascheraInizialmente avevo previsto di utilizzare questo spazio come un contenitore del testo, (ma dopo diverse prove ho deciso, per problemi di leggibilità, di spostarlo in basso, come un tradizionale sottotitolo). La percezione del colore influenza, in parte il comportamento umano. L'effetto che un colore può

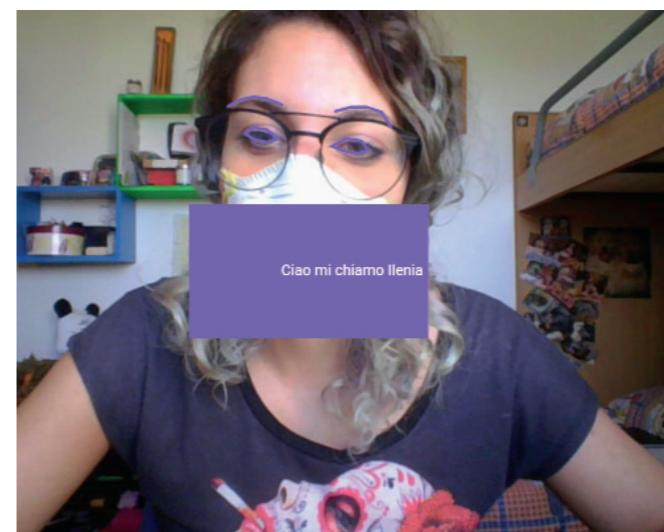


FaceApi + P5Speech + Dandeli (it) Balella Ilenia

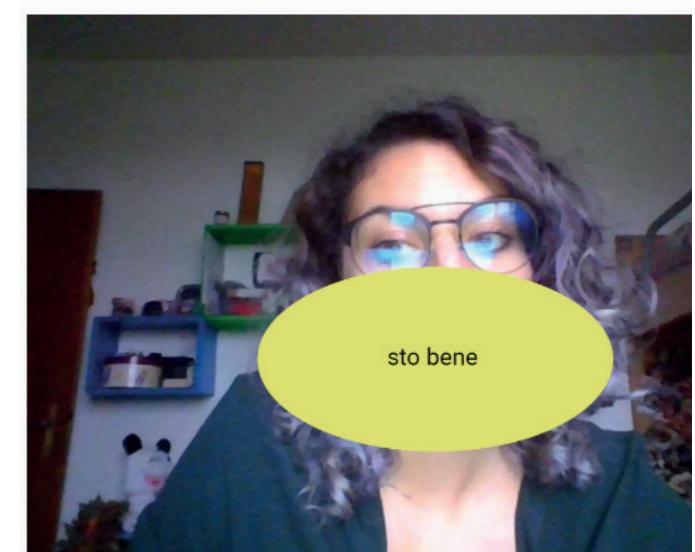


modello caricato: negative

Percentuale sentimento: -70%



FaceApi + P5Speech + Dandeli (it) Balella Ilenia



modello caricato: positive

Percentuale sentimento: 70%



any machine has a list of things you can tell it to do



avere nella nostra psiche è molto personale ed è influenzato anche dalla cultura e società in cui si vive. Perciò è difficile determinare un colore in assoluto che possa essere associato ad una emozione. Il riquadro si modificherà in base alla comunicazione non verbale, cambiando colore e forma a seconda del sentimento analizzato e dall'emozione registrata. Nel prototipo per il momento ai sentimenti (rilevati dalla *sentiment Analysis*) ho associato dei colori arbitrari, per il negativo, il colore rosso, solitamente collegato alla rabbia. Invece per il sentimento positivo il giallo, un colore solare e allegro. A metà tra i due un colore neutro, il blu, che trasmette calma e consente di focalizzare l'attenzione. Ho sviluppato parallelamente i prototipi in lingua inglese sia in italiano, utilizzando in quest'ultimo ho utilizzato la sentiment analysis di dandelion. Ma ho notato che la analisi di quest'ultimo, non è molto affidabile, siccome non comprende esattamente il significato delle frasi, (il suo risultato è quasi sempre neutro). L'alternativa più performante che ho portato avanti è quindi la versione inglese. Ho anche allenato con il teachable machine di google un algoritmo in grado di riconoscere se si indossa una mascherina, per il momento il modello è allenato sulla mia figura, e sarebbe da implementare.

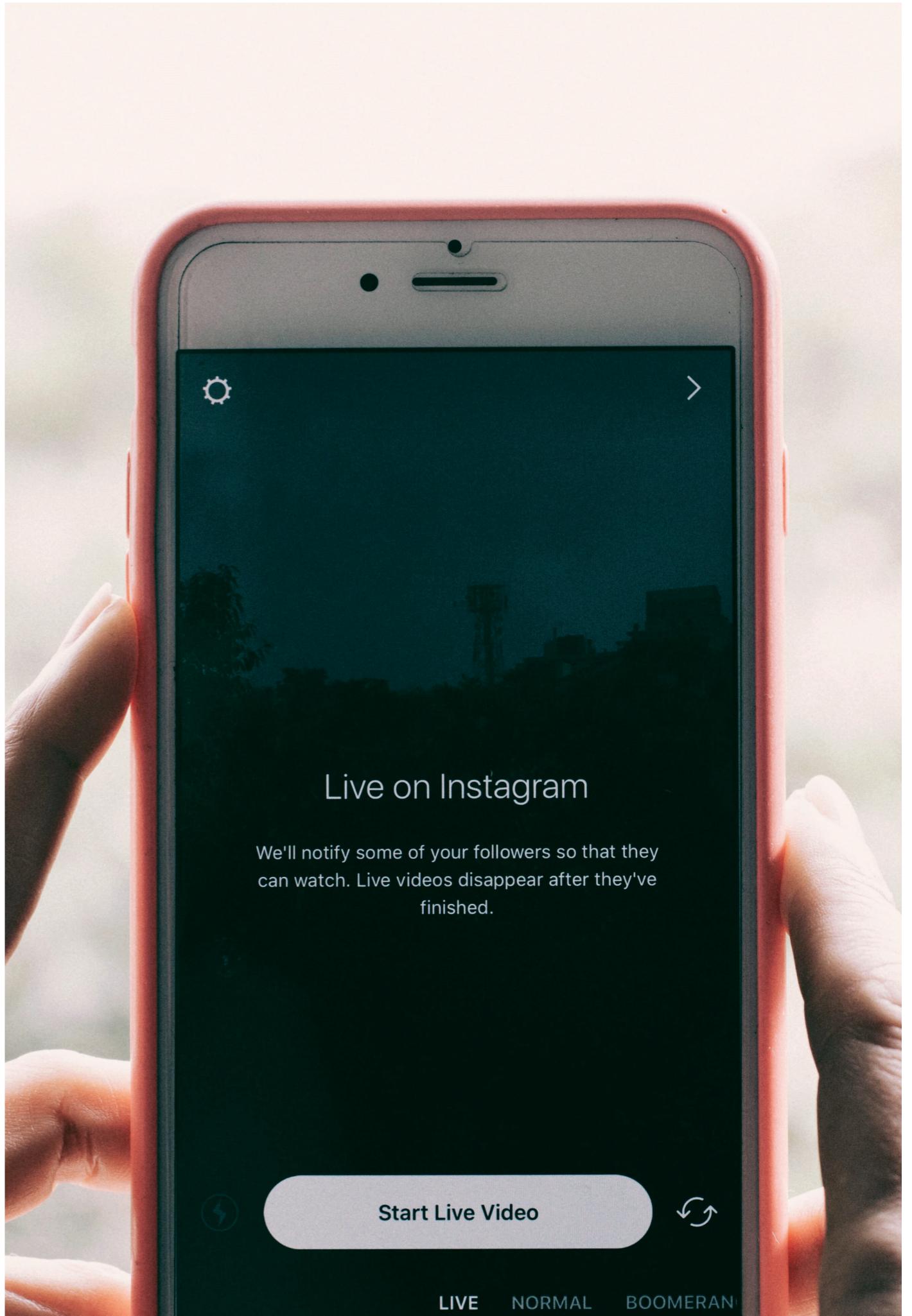
Progetto | che cosa significa

Significa abbattere un muro comunicativo che esclude tutte quelle persone che hanno difficoltà uditive. Le persone sordi infatti, comunicano e comprendono il mondo in modo diverso. Avere un mezzo comunicativo, sempre disponibile nella loro quotidianità, le aiuta ad essere incluse all'interno della società senza sentirsi isolati, in un contesto alterato a causa dell'attuale emergenza.

Progetto | sviluppi futuri

La sottotitolazione è disponibile per video inserzioni, film e serie tv, oppure sulla piattaforma di Youtube con la generazione automatica dei sottotitoli. Ma è fruibile in qualsiasi piattaforma? Quante persone sfruttano le storie o le dirette di instagram per esprimere

a destra
immagine evocativa,
unspalsh.com



i propri pensieri ed esternare la propria identità sociale? Inoltre durante il periodo del lockdown c'è stato un incremento nell'utilizzo utilizzo dei social come strumento di propaganda ed intrattenimento. Tuttavia, non vi è la possibilità di inserire una sottotitolazione automatica (benché in molti video il volto dell'interlocutore è ripreso, ed è visibile il labiale). Ma cosa succede quando il labiale non è inquadrato? Inoltre, se non si ha la possibilità di attivare l'audio perché ci si trova in un luogo pubblico, come si può seguire una diretta? A seguito di questa riflessione si potrebbe pensare ad un plausibile sviluppo futuro. Implementare un filtro instagram che integri in automatico la sottotitolazione nei video sfruttando la libreria di P5 speech. Aumentando la probabilità di raggiungere un pubblico più ampio, includendo anche persone con disabilità uditive.

Cosa succederebbe se si ribaltassero i ruoli, ovvero se si potesse tradurre in parole la lingua dei segni?. L'applicazione potrebbe aiutare entrambe le parti a comunicare ed interagire nella maniera più completa. Si dovrà addestrare un algoritmo che riconosca il linguaggio dei segni, creare dei modelli su cui addestrare la PoseNet.

a destra
immagine evocativa,
unsplash.com



Bibliografia.

Sacks, O. (2014). Vedere voci: Un viaggio nel mondo dei sordi. Milano: Adelphi.

Sitografia.

<https://www.abiliaproteggere.net/2020/05/07/mascherine-trasparenti/>

http://www.unapeda.asso.fr/article.php3?id_article=551

https://it.wikipedia.org/wiki/Comunit%C3%A0_sorda

<https://www.business2community.com/business-innovation/can-googles-translatotron-actually-help-the-accuracy-of-ai-translation-02234708>

<https://ai.googleblog.com/2019/05/introducing-translatotron-end-to-end.html>

<https://cutecircuit.com/soundshirt/>

<https://onezero.medium.com/this-outfit-can-make-deaf-people-feel-music-d0df553efa97>

<https://www.yourtypeface.com/>

<https://www.frizzifrizzi.it/2015/08/01/toyota-sensitive-concert-quella-volta-che-ho-toccato-la-musica/>

<https://www.facebook.com/business/help/1675722002698686?id=603833089963720>

<https://courses.ideate.cmu.edu/60-461s2020/smolinaandrew-cmu-edu/04/28/final-project-slime-language/>

<https://medium.com/@tglaiel/how-to-make-a-voice-activated-led-facemask-diy-guide-b300bc146f38>

<https://github.com/TylerGlaiel/voicemask>