МЕЃУНАРОДЕН СЛАВЈАНСКИ УНИВЕРЗИТЕТ

ФАКУЛТЕТ ЗА **ТЕХНИЧКИ НАУКИ И ИНФОРМАТИКА**
НАСОКА: **КОМПЈУТЕРСКИ СИСТЕМИ И МРЕЖИ**

МАГИСТЕРСКА РАБОТА

ПРИМЕНА НА БИОСЕНЗОРСКИ ИНТЕЛИГЕНТНИ СИСТЕМИ – СЛЕДЕЊЕ НА ОЧИ И АНАЛИЗА НА ГОВОР

Ментор:
проф. д-р Сашо Манасов

Изработил:
Илија Ангеловски

Maj, 2024
Св. Николе

ВОВЕД

Во текот на минатиот век, истражувачите откриja дека начинот на кој звучат нашите гласови може да ни каже многу за нас самите, вклучувајќи ги и нашите емоционални состојби и кои болести можеби ги доживуваме. Гласовната анализа е особено моќна поради нејзината разновидност, неинвазивност и ниска бариера за влез. Создаден е модул за гласовна анализа (audEERING, водечка компанија во развивање на софтвер за говорна анализа од степен на истражување) кој им овозможува на корисниците да додадат говорна анализа во нивниот мултимодален дизајн на научно истражување. Овозможува да се нурнеме во биологијата на нашиот глас за да извлечеме важни сознанија за нашето однесување и психолошка анализа. На што треба да размислиме кога спроведуваме истражување на гласовна анализа?

Гласовната анализа се користи во многу апликации, но нејзината вредност е особено добро докажана во контекст на клинички истражувања. На пример, гласовната анализа може да идентификува биомаркери за Паркинсонова болест, депресија и малигни полипи на грлото. Гласовната анализа сé повеќе се користи во општествените и бихејвиоралните науки каде што може да открие промени во емоционални состојби. Анализата на гласот денес се движи во нови истражувачки полиња како што се: автомобилската индустрија, човечките фактори и науката за перформансите.

Гласовната анализа е моќна алатка која обезбедува увид во човечкото однесување и психологија. Анализирањето на аспектите како тон и ритам помага да се разберат емоционалните состојби и благосостојбата. Тој е разновиден, неинвазивен и има ниска бариера за влез, што го прави привлечен за истражувачите. Гласовната анализа се користи за:

- препознавање емоции;
- следење на здравјето;
- откривање лаги;
- идентификација на говорникот.

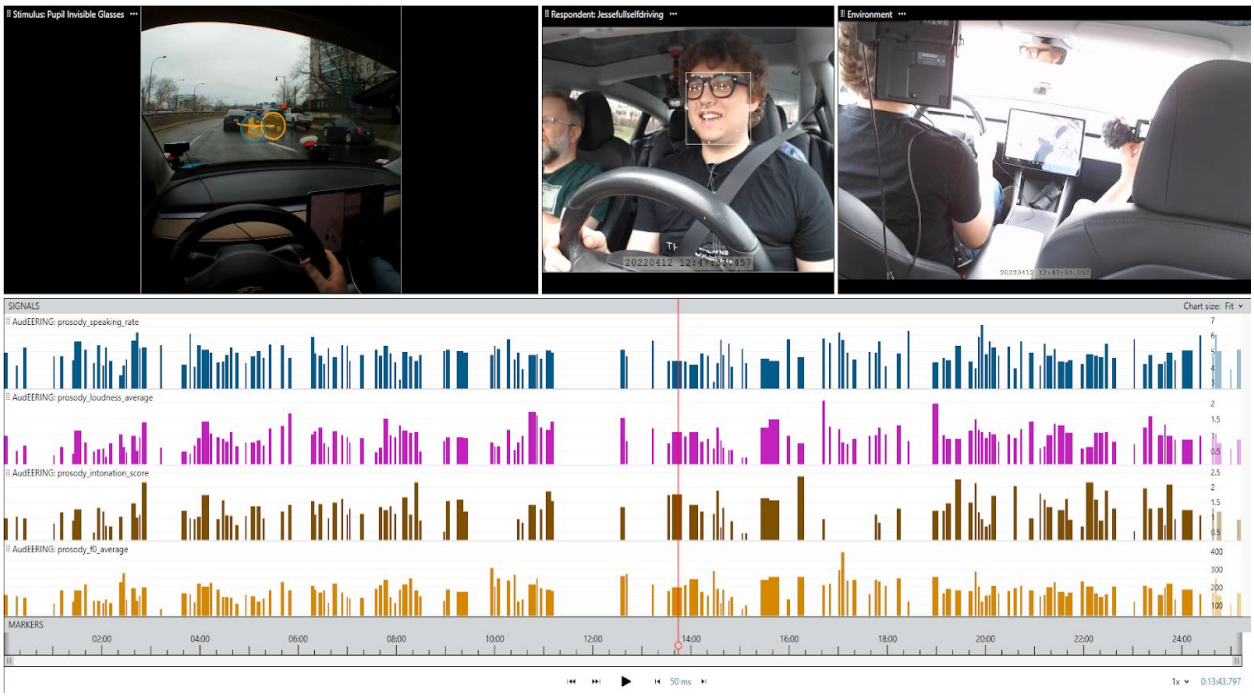
1. Карактеристики на говор – што ние анализираме и интерпретираме

Звукот на нашиот глас е полн со различни карактеристики кои можат да се измерат и анализираат. Во овој дел ќе се фокусирам на три нивои на гласовни карактеристики (прозодија, категории на емоции и димензии на емоции) и што можам да извлечам од секоја од нив.

Овие звучни карактеристики може да се поделат на три аналитички нивои. Што се тие и како се мерат со помош на моделот за гласовна анализа на audEERING во iMotions.

Ниво 1: Прозодија (Prosody) - најосновна метрика за заклучување од гласовна анализа.

Прозодична анализа е класичен пристап за анализа на гласовите. Обично се користи за **клинички** и **лингвистички** истражувања. Прозодиската анализа содржи неколку карактеристики. Овде ќе се фокусирам на четири прозодични карактеристики до кои може да се пристапи во iMotions.



Слика. Пример за визуелизација на податоци за говорна анализа со гласовна анализа во iMotions

- **Висина** (Pitch): Висината се однесува на воочена фреквенција на звукот, конкретно колку се смета дека е висок или низок звукот. Тоа е тесно поврзано со основната фреквенција на звучниот бран, што е вистинска физичка фреквенција на која вибрира звучниот бран. Повисоките фреквенции резултираат со поголем воочен тон, додека пониските фреквенции се перципираат како помал тон.Причина зошто тонот е субјективна мерка се должи на физиологијата на нашето уво. Во нашето уво има мембрана која содржи „влакнести клетки“ кои реагираат на звуците што ги слушаме. Сепак, овие влакна се специјализирани за откривање на специфични звучни фреквенции и тие не се рамномерно претставени. За одреден опсег на фреквенции (мерено во Hz) ние сме многу подобри во откривањето на звуците, едноставно затоа што имаме повеќе влакнести клетки во тој опсег на фреквенција. Ова објаснува зошто опсегот на човечкиот слух е помеѓу 20 Hz до 20 kHz, но обично повеќето возрасни не можат да слушнат звуци повисоки од 15-17 kHz. Колку сме добри во откривањето звуци и како ги перцепираме звуците зависи и од поединецот (второто е добро проучено поле наречена „психоакустика“). Генетските предиспозиции или изложеноста на околината може да влијаат на нашата способност да детектираме и перципираме звуци. **Основната фреквенција** се нарекува “**F0**“ и се мери во херци (Hz). Таа се однесува на бројот на вибрации на гласните набори или циклуси на основната фреквенција што се случуваат во една секунда. Во iMotions, се добива пристап до 4 статистички метрики на F0, кои се пресметуваат преку еден говорен сегмент: **минимална** F0, **просечна** F0, **максимална** F0 и **варијација** на F0. Која метрика е најразбирлива зависи од нашето истражувачко прашање.

Просечна основна фреквенција F0, на пример, може да се користи за да се направи разлика помеѓу машки и женски гласови (85-155 Hz наспроти 165-255 Hz, соодветно). Висок максимум F0 со нормален просечен F0 може да укаже на

спонтани вокални испади со висок тон, како што се изрази на **изненадување** или **одвратност**.

- **Гласност (Loudness):** Гласноста е мерка за тоа колку гласно човек перцепира звук. Тоа е поврзано со амплитудата на звучниот бран, но ги опфаќа физиолошките својства на нашиот слух, кои на експертите им се познати како **психоакустика** (psychoacoustics). Звучен бран со двојно поголема амплитуда кај нас не се доживува како двојно погласен туку, се смета дека е приближно 1,4 пати погласен. Односот помеѓу амплитудата на звучниот бран и воочената гласност не е линеарна, туку **логаритамска**. Понатаму, нашето уво е почувствително на фреквенции во среден опсег (1-3 kHz), кои ги гледаме како погласни од ниските или високите фреквенции со иста амплитуда.

Мерката за гласност во iMotions ги зема предвид овие својства на нашиот слух. Мерката е инспирирана од научниот модел на гласност на Zwicker, кој ја мери гласноста во единицата наречена „**Соун**“ (“Sone”).

Во iMotions, **моделот е поедноставен за работа** со сите видови извори на звук и микрофони и затоа нема калибрација на физичките единици на енергијата на звучните бранови, мерката за гласност е вредност на резултат која се движи од **0,0** (**тишина**) - **1,0** (**максимална можна гласност**) без физичка мерна единица.

Слично на **F0**, **гласноста на еден говорен сегмент** може да се мери како **минимална**, **просечна**, **максимална** или **варијација**. Во клинички студии редовно се пријавува гласност за да се направи разлика помеѓу популација на пациенти.

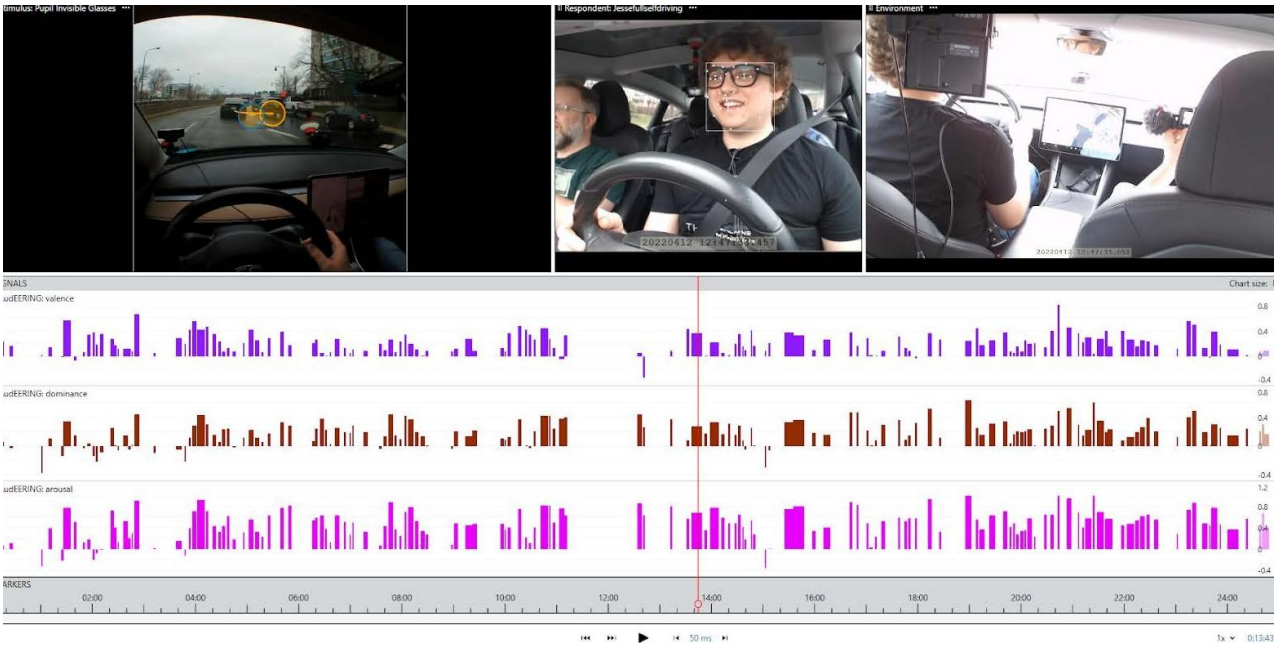
Поради поврзаноста на **гласноста** со **амплитудата** на сигналот, мерката на гласност е под влијание на нивото на снимање на микрофонот (gain) и растојанието на микрофонот и изворот на звук (на пример, лицето што зборува). Ако овие услови варираат во текот на нашето поставување, не треба да ги споредуваме вредностите на гласноста кај овие различни услови, туку во постојани услови, како што е истиот микрофон (идеално слушалки/ревер на фиксно растојание до устата) и исти нивоа на снимање (оневозможување автоматска контрола на нивото на снимање или автоматска контрола на засилување (AGC - Automatic Gain Control)!).

- **Брзина на зборување (Speaking rate)**: Стапката на зборување се однесува на брзината со која зборуваме и се мери како број на слогови во секунда и неговата варијација во еден говорен сегмент. Кога варијацијата на брзината на зборување е мала, тоа покажува дека гласот има стабилно темпо, додека високата варијација укажува на менување на темпото, на пример: забавување на важни исказни делови и забрзување на помалку важни делови. Високата варијација, исто така, може да укаже на присуство на исполнети паузи и двоумење (како „**ух**“), поради **гласно размислување** или **поголемо когнитивно оптоварување**. Вредности на просечна стапка на зборување од **3-5** се нормални, **под 3** побавен говор и **над 5** побрз говор. Ако има повеќе паузи во говорниот сегмент, измерената просечна вредност може да биде помала од вистинската воочена стапка на говорење.
- **Интонација (Intonation)**: Интонација е **мерка за порастот и падот на гласот** (на пример: мерка за тоа колку јачината се менува во еден говорен сегмент). Корисно е за мерење колку **монотоно** (**ниска интонативна вредност**) или **живо** (**висока интонативна вредност**) зборува човекот. Структурата на реченицата не е земена во предвид, односно не се врши целосна анализа на интонациските обрасци. Интонациските резултати се изведени од стандардното отстапување на висината и имаат типичен опсег од 0,4 – 1,6.
    - Вредностите под **0,4** укажуваат на **монотон** говор.
    - Вредностите над **1,6** укажуваат на **живо зборување**.
    - Вредностите од **1,0** укажуваат на **нормална**, **просечна интонација**.

Ниво 2: **Емоционални димензии** – доловување на емоционални изрази во гласот.

Анализата на емоционалната димензија е особено корисна во општествени и бихејвиорални истражувања кои се заинтересирани за **карактеристики на**

личноста и **перформансите**. Постојат три емотивни димензии и тие обично се визуелизираат во **3Д простор** со биполарни оски (слика).



Слика 2. Пример за визуелизација на податоци за говорна анализа во iMotions, покажувајќи карактеристики на емоционална димензија

- **Активација** (возбудување - arousal): Активирањето го покажува **нивото на возбуда на звукот на гласот** и се мери во опсег од **-1,0** (**ниско**) до **1,0** (**високо**), каде што **0,0** е "**неутрално**", т.е. нормално ниво на возбуда што се очекува од неутрален глас кој чита текст- параграф. Активирањето може да се користи за да се заклучи дали некој настан имал **смирувачки** (calming/ soothing) или **возбудувачки** (exciting/agitating) ефект врз некоја личност. Општо земено, ниските резултати за активирање може да се толкуваат дека укажуваат на **тага**, **релаксација**, **замор**, **депресија** или **задоволство**, додека високите резултати за активирање обично укажуваат на **радост**, **страв**, **бес** или **возбуда**. Понатаму, активирањето е индикатор за лична релевантност (personal relevance). Зголеменото активирање покажува дека:
    - личноста е емотивно поврзана со темата или стимулот за кој зборува;

- има интерес за темата;
- има одредена итност да дознае информации преку говорот на другите во разговорот.

- **Доминација** (Dominance): Доминацијата го означува нивото на доминација на звукот на гласот и се мери во опсег од **-1,0** (**ниско**) до **1,0** (**високо**), каде што **0,0** е "**неутрално**", т.е. умерено ниво на доминација што би се очекувало во неутрален глас кој чита текст параграф. **Доминацијата** може да се користи за да се заклучи **колку говорникот се чувствува покорен** или **контролиран**. Како примери, ниската доминација укажува на состојба на **страв**, додека високата доминација укажува на состојба на **гордост**.
- **Валентност** (Valence): Валентноста го означува степенот на **позитивност** или **негативност** присутен во гласовниот израз. Се мери во опсег од **позитивен** (**1,0**) до **негативен** (**-1,0**), каде што **0,0** е **неутрален**. Валентноста може да се користи за да се заклучи дали настанот имал **непријатен/негативен** (unpleasant/negative) или **пријатен/позитивен** (pleasant/positive) ефект врз некоја личност. Генерално, негативната валентност укажува на состојба на **лутина**, **тага**, **страв**, или **досада**. Спротивно на тоа, позитивната валентност укажува на состојба на: **релаксација**, **задоволство**, **среќа** и **возбуда** (excitement).

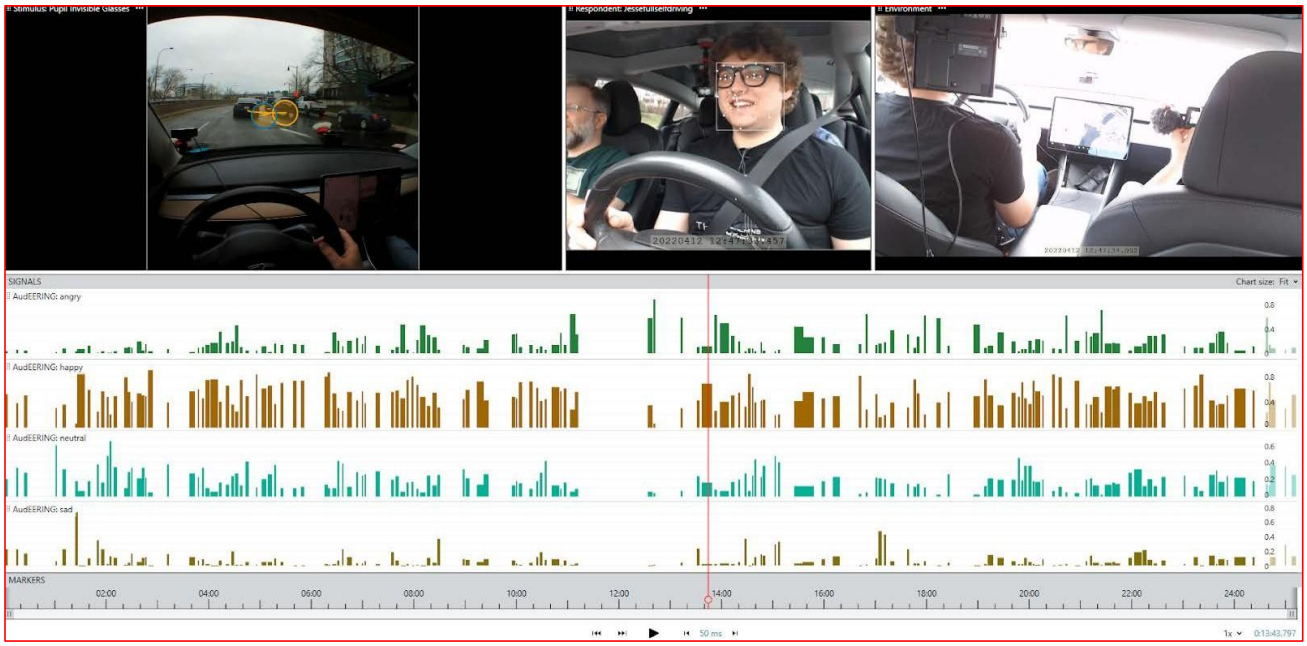
Ниво 3: Категории за изразување емоции (Emotion expression categories) – **доловување на изразување на емоции во 4 основни категории**.

Овие метрики на високо ниво (high-level metrics) се особено корисни за студии кои ги:

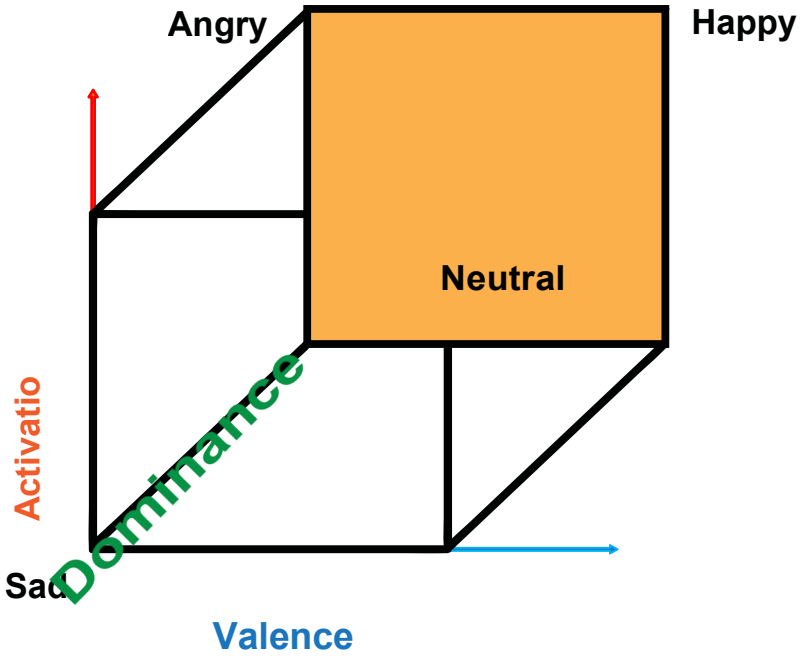
- **проценуваат перформансите на луѓето**;
- **гласните емоционални реакции на настани**.



Постојат **4 категории на изразување емоции** достапни во iMotions, кои се изведени од гласот: **среќен**, **неутрален**, **тажен** и **лут** (happy, neutral, sad and angry). Нивните резултати се движат од **0** (**најниска**) до **1** (**највисока**), а **оценките во секој гласовен сегмент** (повеќе за гласовни сегменти во текстот што следи) ќе имаат збир од 1. **Емоција со највисока оценка** (т.е. **поблиску до 1**), го претставува **најдоминантниот израз на емоции** во текот на тој гласовен сегмент. Ако две категории на емоции добијат сличен резултат, моделот не е сигурен кој израз е присутен.



Слика. Пример за визуелизација на податоци за говорна анализа во iMotions, која прикажува податоци за категориите на емоции



Слика. Димензии на емоции и категории на емоции често се визуелизираат во 3Д простор со биполарни оски, што визуелизира како резултатите од димензијата на емоциите корелираат со оценките на емоциите

2. Видови истражувачки студии со користење на гласовна анализа

Гласовна анализа (Voice analysis) се користи во многу различни видови истражувања и можеме да размислуваме за користење на гласовна анализа врз основа на видот на полето за истражување или типот на поставеното истражување. Во продолжение прво давам примери за тоа како гласовната анализа се применува во три различни истражувачки области, а потоа и како гласовната анализа се користи со различни истражувачки дизајни.

Постојат три главни истражувачки области кои користат гласовна анализа:

2.1 Клиничко истражување (Clinical research):

Клиничкото истражување со децении користеше гласовна анализа за да направи разлика помеѓу клиничката популација. На пример, студиите идентифи-

куваа гласовни биомаркери за дизартричен говор, депресија, Алцхајмерова болест и Паркинсонова болест (Narendra и Alkul, 2018; Bocklet et al., 2013; Taguchi et al., 2018 Han et al., 2018; Meilan et al., 2013). Во клиничките истражувања, гласовната анализа се користи и за да се процени ефикасноста на третманот. На пример, со студии кои ја оценуваат ефикасноста на третманот за социјална анксиозност со помош на гласовна анализа (Laukka et al., 2008).

Во клиничкото истражување, гласовната анализа обично се објавува како дел од високо контролирани, лабораториски истражувачки студии каде што учесникот извршува говорна задача додека е соочен со компјутер или како дел од клиничко интервју.

2.2 Социјално и бихејвиорално (за однесувањето) истражување:

Гласовната анализа е многу популарна во областа на социјалната и бихејвиоралната наука подобро да разбереме:

- **како нашиот глас ја одразува нашата личност**;
- **како влијае на луѓето околу нас;**
- **како различни фактори влијаат на тоа како ние перципираме други гласови**.

На пример, познати студии процениле како менструалните периоди ја менуваат желбата на жените за машки гласовни карактеристики (Puts, 2005), додека други студии проценувале како промената на звукот на нашиот глас влијае на нашата самоперцепција (Stel et al., 2011). **Гласовна анализа е користена и за предвидување на исходот од изборите** (Banai et al., 2017).

Исто како и клиничкото истражување, општествените и бихејвиоралните науки често пати ги вршат своите истражувања за гласовна анализа во контролирани лабораториски поставувања. Сепак, повообичаено е ова поле на истражување **да користи однапред снимени видео или аудио датотеки за да ги анализира гласовите од „реалниот свет“**.

2.3 Маркетинг и бизнис истражувања:

Маркетингот и деловното (бизнис) истражување применуваат гласовна анализа за да оптимизираат неколку аспекти од нивната клучна функција, вклучувајќи ги нивните стратегии за комуникација и продажба. Гласовната анализа се користи комерцијално во центри за повици, UX дизајн и истражување на пазарот, како што е тестирање на производи.

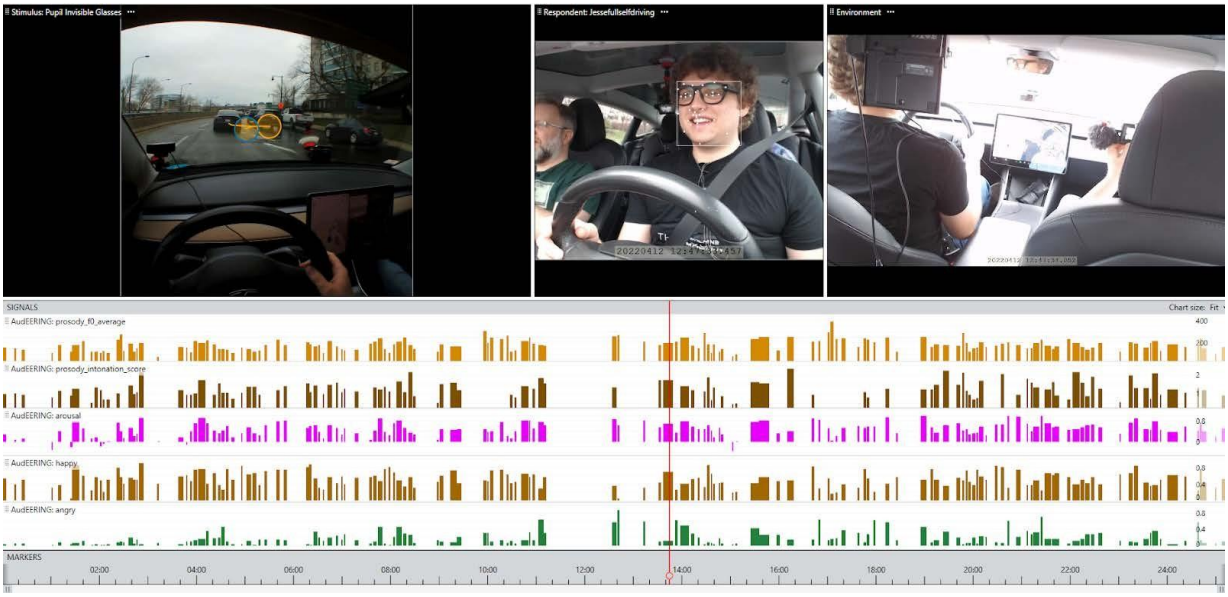
2.4 Типови на дизајн на студии

Гласовната анализа е особено моќна поради нејзината разновидност во собирањето податоци. Затоа е вообичаено да се види гласовна анализа која се користи во многу различни типови дизајни на студии. Некои од најчестите дизајни на студии кои применуваат гласовна анализа се следниве:

- **Тестови за гласно размислување** (Think aloud tests): Тестовите за размислување гласно се однесуваат на дизајн на студија каде што учесниците се упатени да го кажат она што им е на ум додека извршуваат задача. Оваа стратегија често се гледа во истражување на корисничко искуство. Пример истражување (Fan et al., 2019; Fan et al., 2021).
- **Дијадски интеракции** (Dyadic interactions): Дијадите се однесуваат на две лица кои комуницираат. Истражувањето на дијади е различно од интервјуа со интервјуер и соговорник, бидејќи тие обично се случуваат како природни разговори помеѓу двајца колеги, пријатели или панелисти. На пример, гласовната анализа може да се примени за да се процени емоционалната состојба на дијадите за време на тимска задача (Neubauer et al., 2017), или да се истражи тонот и модификацијата на гласот помеѓу луѓе со низок и со висок статус (Gregory and Webster, 1996).
- **Презентации:** Презентирање со привлечен глас станува сè поголема побарувачка, а гласовната анализа веќе одигрува улога во разликувањето помеѓу многу привлечни и слабо ангажирани говорници.

Пример: студија која ги споредува гласовните карактеристики на Стив Џобс за време на презентации со просечен говорник (**Niebuhr et al., 2016 - Comparing the voice features of Steve Jobs during presentations**). Во тие линии има смисла дека гласовната анализа се применува за оптимизација на перформансите и коучинг за јавно говорење.

- **Интервјуа:** Интервјуата се редовни активности за работни места и здравствените установи. Клиничките интервјуа се особено добро проучени со гласовна анализа, на пример во депресија (Scherer et al, 2014) и коронарна болест (Schucker et al., 1977).
- **Автомобилизам**: Иако ова е сè уште ново поле, гласовната анализа дава големо ветување за автомобилско истражување. Главна причина за овој потенцијал е што **гласовна анализа е неинвазивен сензор** што може да се снима во автомобилот без да се бара од лицето да стави дополнителен хардвер (иако можеби се размислува за микрофон за највисок квалитет на податоци во зависност од бучавата од околината).



Слика: Пример за гласовна анализа во автомобилска студија. Јутјубери на Now You Know го посетија iMotions за да тестираат како различни биосензори го доловуваат различното искуство на возење на обично или самоуправувачко возило Tesla. Гласовната анализа овозможи, заедно со GSR, очила за следење на очите и ЕКГ, да се открие кои моменти се најзненадувачки, стресни и когнитивно напорни за возачот.

Гласовна анализа се користи за:

- откривање на замор;
- употреба на супстанции и може да ја промени играта на автомобилските компании кои се стремат автоматски да ја идентификуваат когнитивната состојба на возачот. Се реализираат внатрешни студии за да се покаже изводливоста на овој пристап.

Заклучок: Гласовната анализа се користи во различни истражувачки полиња, вклучувајќи клиничко истражување, социјални и бихејвиорални истражувања и маркетинг и деловни истражувања. Гласовната анализа може да се примени на збирки податоци собрани со различни дизајни на студии, вклучувајќи тестови со гласно размислување, дијадични интеракции, презентации и интервјуа.



3. Процесирање (обработка) на сигнали

AudEERING е германска компанија за вештачка интелигенција основана во 2012 година како дел од академски кругови каде што основачите ја развиле сега широко користената програма за анализа на глас со отворен код **openSMILE**. AudEERING ги изгради своите последователни производи за анализа на говор и

емоции врз основа на openSMILE и сега нуди пософистицирани и елаборирани карактеристики за анализа од она што е достапно преку **openSMILE**, вклучувајќи го нивниот комплет за развој на софтвер **devAIce®** (SDK) кој е производ што се користи кога се прави гласовна анализа со iMotions. До денес, audEERING има 11 јавно финансирани истражувачки проекти, а нивните производи се користат од широк спектар на клиенти кои опфаќаат академски и комерцијални апликации.

Ако се прави гласовна анализа во iMotions, се користи производот devAIce® на audEERING, кој обезбедува говорна прозодија и стандардно извлекување на сет на функции, како и анализа на димензии и категории на изразување на гласовни емоции.

Како се анализираат податоци: постпроцесирање

Можеме да снимаме аудио во iMotions (на пример од камера на сцена или камера на испитаник за време на студија базирана на екран), или може да внесеме во iMotions аудио снимка собрана со друг уред (на пример од паметен телефон). Откако ќе ја имаме нашата аудио снимка во iMotions, можеме да го изберете копчето „**Процесирање на глас**“ (“**Post-process voice**”) на дното на екранот. Во дијалогот што ќе се појави, можеме да избереме кој аудио извор да го користиме, а потоа да ја започнеме гласовната анализа, која ќе работи во позадина. Во зависност од должината на нашите снимки, обработката на гласовната анализа може да потрае неколку минути. Гласовната анализа ќе се реализира со помош на devAIce® со audEERING користејќи ги поставките што сме ги избрале во глобалните поставки на iMotions.

Откако гласовната анализа ќе ја заврши својата обработка, можеме да го отвориме прозорецот за повторување (the replay window) за еден од нашите испитаници и ќе можеме да ги визуелизираме гласовните сигнали со нивно избирање во левиот панел. Гласовните сигнали автоматски се синхронизираат и со аудио/видео снимањето и со сите други сигнали што можеби сме ги собрале за време на реализирање на снимањето.

iMotions локално го интегрира devAIce® SDK, што значи дека нашите аудио податоци се анализираат целосно на нашиот компјутер и никогаш не се

поставуваат податоци на сервер. Ова ни го олеснува усогласувањето со локално важечките прописи за приватност.

а) Сегмент на говор (Speech Segment): откривање на гласовна активност

Компонентата Speech Segment е специјално дизајнирана да открие присуство на човечки глас во аудио пренос. Оваа детекција е цврста (робусна) за бучава и е независна од нивото на јачината на гласот (што значи дека слабите гласови може точно да се детектираат во присуство на погласни звуци во позадина).

Времетраењето на говорниот сегмент ја одредува должината на секој говорен сегмент. Секој говорен сегмент се анализира за **прозодија** (prosody), димензии на емоции (emotion dimensions) и категории на емоции (emotion categories). Времетраењето на Speech Segment (говорен сегмент) може рачно да се одреди, што значи дека нашата гласовна аналитика може да рефлектира периоди на единствени временски периоди во зависност од нашето проучување. Треба да се има предвид дека можеме да избереме само едно времетраење на говорниот сегмент и тоа ќе остане исто за целиот експеримент. Препорачаното минимално времетраење на секој сегмент од говорниот сегмент е **1,5 секунди** и се препорачува да има **најмалку 30 секунди говор** **по говорник** и **стимул**.



б)Сетови од функции Prosody и openSMILE

devAIce® дава неколку прозодични карактеристики кои најчесто се користат во истражувањето:

- **основна фреквенција** (fundamental frequency) (Hz): просечна вредност по говорен сегмент (average value per speech segment);
- **интонација** (intonation) (оценка 0-1, низок до висок): просечна варијација на основната фреквенција по говорен сегмент (average variation of fundamental frequency per speech segment);
- **гласност** (loudness) (0 -1 резултат, низок/ тивко до високо/гласно), и
- **стапка на зборување** (speaking rate) (број на (псевдо - слогови во секунда).

Дополнително, devAIce® дава пристап до два стандардни сетови на функции на openSMILE: **GeMAPS** (Eyben et. al, 2015) и **ComParE 2016** (Schuller et al., 2016). ComParE 2016 е најголем сет на функции со над 6K карактеристики (Schuller et al., 2016), и двата сета на функции се високо цитирани во академски списанија (со над 2600 цитати) (Eyben, 2016; Eyben et al., 2010). Овие сетови на функции претежно се користат од академици кои сакаат да градат свои модели врз основа на големи комплети за екстракција на аудио функции. За повеќето истражувачи, прозодичните карактеристики директно достапни преку devAIce® (и видливи во датотеките за повторување и експортирање на iMotions) се доволни за нивно истражување.

b) Технологијата devAIce® на audEERING за анализа на изразување на гласовни емоции

The devAIce® технологијата/пакет ја користи најновата технологија за вештачка интелигенција на audEERING за да постигне изразување на гласни емоции на **димензии на активирање**, **валентност** и **доминација**, како и избрани категории на изразување емоции. Овој модел е изграден на голема длабока невронска мрежа со милиони параметри, која ја пресметува димензијата на емоциите и

резултатите од категоријата како континуирани вредности врз основа на аудио влезот. Моделот е обучен за поврзување на аудио -влезовите и димензиите и категориите на емоции од голем број примероци за обука означени од човечки експерти. Не може точно да се каже кои гласовни карактеристики ги користи моделот во своите пресметување на димензиите на емоциите или категориите на емоции, но моделот е енергично тестиран на емоционални искази со глумење и без глумење со многу различни звучници, микрофони и акустични средини. Врз основа на овие тестови, моделот кој стои зад технологијата devAIce постигнува точност од 65% за откривање емоции (**среќен**, **тажен**, **неутрален**, **лут**) и **коефициент на корелација на усогласеност** од **0,75**, **0,64** и **0,63** за **возбуда**, **валентност** и **доминација**, соодветно, што е на исто ниво како и точноста на еден човек експерт оценувач во споредба со просекот на експертски оценувачи.

Моделот **Emotion Large** известува за резултати врз основа на секој сегмент од Speech Segment и користи изговорени зборови како дополнение на гласовните карактеристики за анализа на резултатите од емоциите. Моделот за гласовна анализа го зема предвид само англискиот јазик и кога со тоа малку се подобрува точноста на гласовната анализа, особено валентниот резултат. Не е можно да се оневозможи анализата на говорниот јазик на моделот, но необјавените тестирања од audEERING и други групи откриваат дека точноста на моделот Emotion Large е генерално слична помеѓу англиски и неанглиски говорници.

c) Атрибути на говорникот (Speaker attributes): воочување на пол и возраст

devAIce® дава атрибут за пол и возраст врз основа на гласовните карактеристики и длабоката вештачка интелигенција на audEERING, која видела над 2 милиони говорни сегменти. Атрибутот род не е поврзан со родовиот идентитет на говорникот. Тоа е чисто проценка на воочениот биолошки родов атрибут на гласот, т.е. ако гласот звучи како машки или женски глас.

devAIce® може да предвиди дали говорникот е женско, машко или дете (постои дополнителна можност за предвидување возраст, што е опишано подолу). За секој говорник, моделот devAIce дава три излезни метрики (машки, женски и

деца) со ниво на доверба од **0-1**. Излезот со највисоко ниво на доверба (поблиску до 1) е „**победник**“ а со тоа и заклученото предвидување. На пример, ако моделот дава ниво на доверба од 0,7 за маж, 0,2 за женски и 0,1 за дете, финалето предвидувањето е дека говорникот е маж.

Атрибутот за пол може да се пресмета на еден од двата начини:

1. **Како атрибут заснован на целокупната говорна снимка**. Ова дава најголема прецизност, ако гласовното снимање содржи еден говорник.
2. **Како атрибут заснован на секој сегмент од говорниот сегмент**. Ова дава континуирани податоци и интервали на доверба, но ја загрозува точноста..

Карактеристика за возраст (The age feature) ја проценува предвидената возраст на говорникот во години и има просечна апсолутна грешка од 9,94 години кога се тестирани на реални чисти и бучни податоци за говор. Истражување направено за **способноста на човекот** да предвидува возраст врз основа на гласовни извештаи со средна апсолутна грешка од 10,34 години. Ова значи дека технологијата за уреди на audEERING е подобра во предвидувањето на возраста од глас отколку човекот.

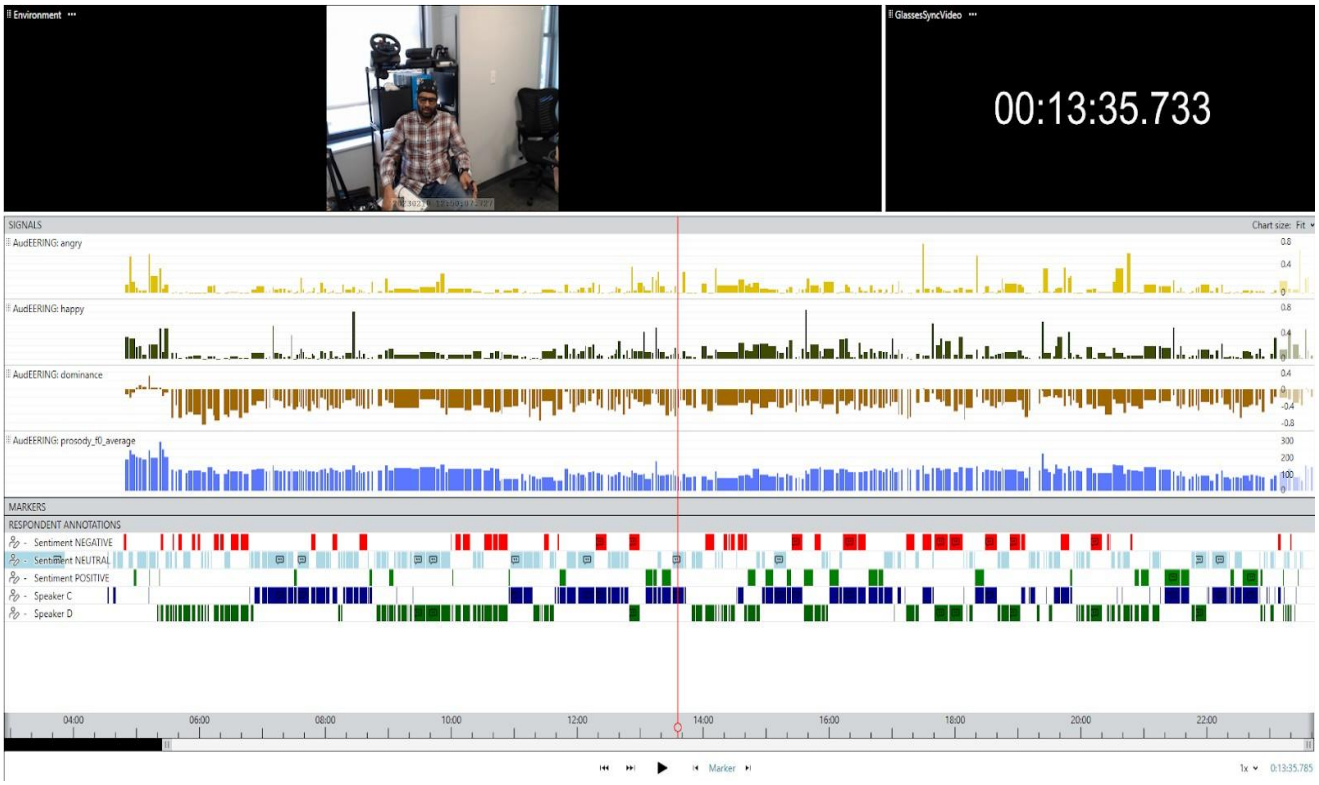
d) Анализа на повеќе говорници (Multi – speaker analysis)

Во многу случаи, студиите вклучуваат само еден говорник, меѓутоа, некои дизајни на студии вклучуваат разговор помеѓу двајца или повеќе луѓе. Ако одредена студија спаѓа во втората група, тогаш важно е да се има предвид дека devAIce® во моментов не може да прави разлика помеѓу повеќе говорници, што значи дека гласовната анализа ќе ги просекува гласовните карактеристики кај сите говорници. Овој предизвик може да се надмине со користење на една или двете карактеристики на јадрото iMotions:

1. **Обележувачи на настани и автоматски прибелешки** (Event markers and auto- annotations): може рачно да се означи кога секој говорник почнува и престанува да зборува и да се креираат автоматски

прибелешки за овие периоди (на пример, говорник А и говорник Б). При експортирање, може да се пристапи до гласовна аналитика подредена според секоја прибелешка и автоматски ќе се одвојат податоците според конкретен говорник.

2. **Способноста за говор во текст преку AssemblyAI** (The Speech-To-Text ability through AssemblyAI): Со користење лиценца за AssemblyAI, може да се искористи нејзината способност за говор во текст за да се одделат помеѓу говорници што може да се визуелизира во прибелешки (како говорник А, говорник Б, и.т.н.). Слично со користењето на маркери за настани и автоматски прибелешки, сега може да се пристапи до гласовни податоци за специфицирани говорници во експортирањето.



Слика. Пример за тоа како може да се користат забелешки и/или функции од говор во текст да се направи разлика помеѓу повеќе говорници за истражување на гласовна анализа

4. Формат на податоци и како да се интерпретираат гласовни карактеристики

а) Вид на податоци до кои се има пристап

Се добива пристап до обработени податоци што ги дава моделот devAIce® врз основа на негова **анализа на прозодија**, **анализа на димензија на емоции** и **анализа на категорија на емоции**. Овие податоци се добиваат за секој говорен сегмент, а исто така може да се пристапи до податоци од функциите openSMILE по говорен сегмент. Необработени внатрешни податоци од моделот не се достапни.

iMotions моментално обезбедува резиме на сите гласовни податоци и ги организира според забелешките, доколку тие биле користени. iMotions обезбедува и специфичен notebook **за да се анализира која емоција е најзастапена во текот на секој говорен сегмент**. The notebook **ја црта емоцијата со највисока оценка за секој говорен сегмент** за да овозможи лесен визуелен приказ за тоа како емоционалниот тон на гласот се менува со текот на времето.

б)Разлика помеѓу димензиите на емоциите и категориите на емоции, и како тоа влијае на нашите интерпретации

Димензиите на емоциите опфаќаат три аспекти на нашиот глас

- **колку активирачко звучиме** (how activated we sound);
- **колку доминантно звучиме;**
- **дали звукот е позитивен или негативен**.

Димензиите на емоциите се независни оценки и можно е говорникот да постигне високи или ниски оценки на сите нив. Спротивно на тоа, резултатите од емоциите содржат четири емоции:

- **колку звучиме среќно**;
- **колку звучиме луто**;
- **колку неутрално звучиме;**
- **колку тажно звучиме**.

Секој говорен сегмент добива оценка за сите 4 емоции, што вкупно изнесува **1**. Емоцијата со највисока оценка се смета за вистинска емоција изразена во гласот. Така, едно лице може да има глас кој добива 30% оценка за тага, 10% резултат за среќа, 15% неутрален резултат и 45% резултат лутина (гнев). Во овој случај, предвидената емоција изразена во конкретниот говорен сегмент е гнев. Тоа значи дека кога ги анализираме оценките за димензијата на емоциите, можеме да ги окарактеризираме и како доминантни, возбудени и негативни, но со оценките за емоции ги карактеризираме само како лути. Оценките за димензијата на емоциите корелираат со оценките на емоциите. Така, се очекува дека личноста со силно активиран, доминантен и негативен глас е категоризирана како оценка „**лута**“ за емоции. Во спротивно на тоа, ако лицето има ниски резултати за активирачко и доминација, но сепак со негативна валентност, нивниот глас би бил категоризиран како изразување „**тажна**“ емоција.

б)Што можеме да научиме од карактеристиките на прозодичен глас

Карактеристиките на прозодичниот глас се состојат од богат увид во **живоста** (liveliness) и **квалитетот** на гласот. Истражувачите користат прозодична гласовна анализа за да ги идентификуваат биомаркерите на пациентите со големо депресивно нарушување и Паркинсонова болест.

5. Препораки за собирање податоци

Големина на примерок и времетраење на снимање

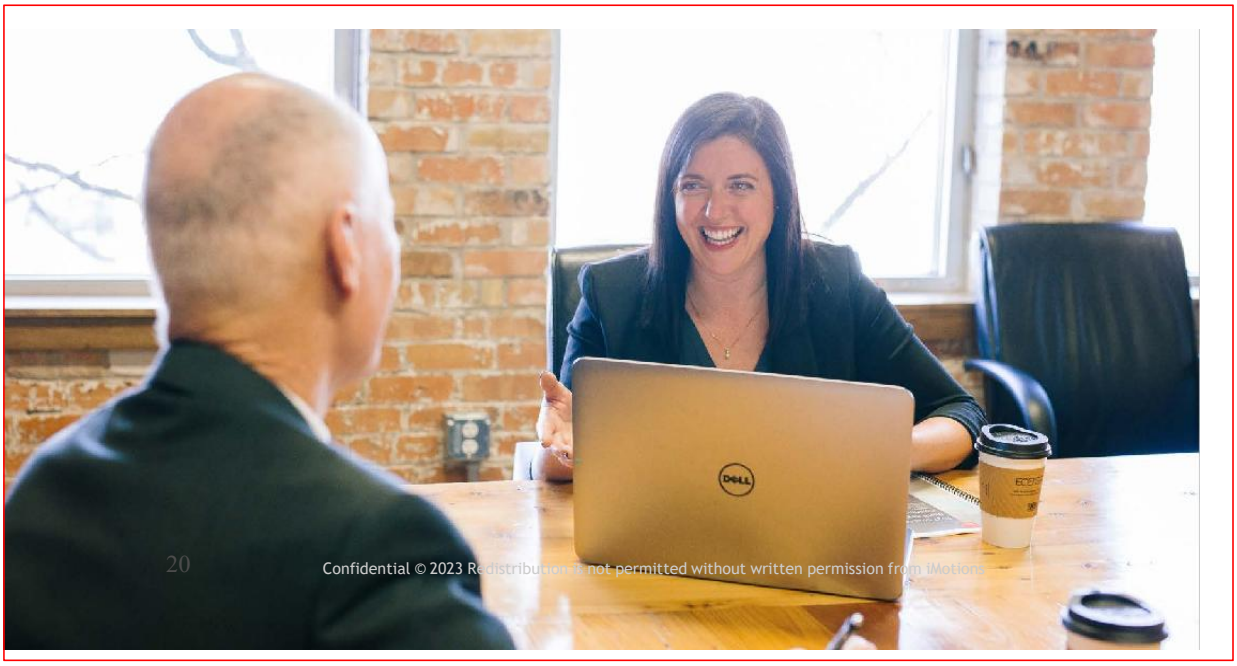
За да се постигнат најсигурни резултати, се препорачува дизајнот на студијата да ги следи упатствата:

- Секој говорен сегмент е во должина од **1,5-10** секунди.
- Најмалку **30** секунди и идеално **60** секунди говор, по лице по стимул.
- Големина на примерок од **20-40** говорници од кои може да се собираат податоци. Оваа препорака ќе зависи од целта на студијата и презентираните стимули.

6. Каква опрема е потребна

Единствената опрема што треба да ја земеме во предвид конкретно за гласовна анализа е изборот на микрофонот.

За повеќето истражувачи, внатрешните микрофони на сите нови уреди од приближно 2020 година и подоцна, вклучувајќи ги лаптопи и десктоп- компјутери, имаат доволно внатрешни микрофони за квалитетна гласовна анализа. Устата на говорникот треба да биде оддалечена околу 10-50 cm од микрофонот. Ако учесникот е подалеку од 50 cm од компјутерот/внатрешниот микрофон, тогаш треба да се размисли за користење на надворешен микрофон за собирање податоци.



Надворешните микрофони се безбедна опција и поголем број на понови микрофони и слушалки со микрофон кои имаат добри резултати. Треба да се погрижиме да избереме микрофон поврзан со USB. audEERING ги фаворизира микрофоните со ревер (Lapel microphones). Се препорачува да не се користи микрофон на надворешни веб-камери или евтини микрофони поврзани со USB, бидејќи тие ќе резултираат со многу понизок квалитет на податоците. Микрофони

од студио обично не се потребни, но се охрабруваат за истражувачи кои се многу сериозни во врска со квалитетот на гласовните податоци.

audEERING ги препорачува следните карактеристики на микрофоните што се користат за гласовна анализа:

- Микрофонот треба да може да собере минимална брзина на примерок од 16 kHz со 16-битна квантизација.
- Идеално, снимањето е во некомпресиран формат. Ако е компресирано, треба да се провере дали е MP3 со брзина од најмалку 64 kbit/s.
- Фреквентниот опсег на микрофонот треба да биде 50 - 10.000 Hz или поширок.
- Генерално се претпочита да се користат **омнинасочни микрофони** (omni-directional microphones), и **кардиоидни микрофони** (cardioid microphones) сетирани со поголем шум во позадина.

Неопходно е секогаш да се користи ист микрофонот при целото собирање податоци за дадена студија за да се обезбеди точна споредба помеѓу учесниците, бидејќи квалитетот на звукот ќе влијае на квалитетот на гласовната аналитика.

Аудио- придобивки

Поради поврзаноста на **гласноста** со **амплитудата** на сигналот, мерката на гласност е под влијание на микрофонско ниво на снимање (добивка) и растојанието на микрофонот и изворот на звук (на пример, лицето што зборува). Ако овие услови се менуваат во текот на нашето поставување, не треба да ги споредуваме вредностите на гласноста во овие различни услови, но повеќе во постојани услови, како што е ист микрофон (идеално слушалки/ревер на фиксно растојание до устата) и исти нивои на снимање (да се оневозможи автоматска контрола на нивото на снимање или функции за автоматска контрола на засилување), (AGC - automatic gain control).

Соодветна средина за снимање

Секогаш се препорачува собирање гласовни снимки во средини со минимален шум во позаднина. За најточна и најчувствителна гласовна аналитика, гласовните податоци треба да се снимаат во добро изолирани простории со мини-

мален шум во позадина, иако ова размислување е најважно за истражувачите кои идентификуваат суптилни разлики во гласовните карактеристики помеѓу експерименталните групи. Бидејќи компонентата “Speech Segment“ е специјализирана за откривање човечки глас, истражувачите не мора да обезбедат специјализирани локации или опрема за да добијат точни резултати.

Заклучок:

- Гласовната анализа е неинвазивен сензор за разбирање на човековото однесување, биомаркери и емоции.
- Гласовната анализа во голема мера се користи во клинички и лингвистички истражувања и најчесто се применува во компании кои ги проучуваат перформансите (на пример, центри за повици) и корисничко искуство.
- Гласовната анализа дава три слоја на метрика: **прозодија** (најгрануларна), **димензионална анализа** и **категории на емоции**.

7. Воведување на говор во текст и анализа на валентност

Интеграцијата на вештачката интелигенција (AI - artificial intelligence) направи револуција во различни индустрии, а денес ја трансформира областа на истражувањето на човековото однесување. iMotions во 2023 година го интегрира API за софистицираното препознавање и анализа на говор на AssemblyAI во сопствената платформа iMotions Lab, за да може да им понуди анализа од говор во текст на своите корисници.

Оваа интеграција означува значајна пресвртница во научните и истражувачки напори за собирање податоци и анализа, овозможувајќи им на корисниците да отклучат подлабоки сознанија од говорниот јазик и да ги поттикнат истражувачите и аналитичарите со зголемени способности за разбирање и толкување на човечката комуникација.



Слика. Анализирање и сегментирање на аудио и видео- датотеки, со препознавање говор, анализа на валентност и автоматска транскрипција

Импортирање, транскрипција и анализа

Новата функција за анализа на говор во текст им овозможува на корисниците да импортираат видеа или аудио-датотеки и преку API на AssemblyAI, аудиото автоматски да се транскрибира и анализира. Анализата на звукот вклучува откривање на број на говорници, дискурзивно откривање на валентност (употреба на позитивни, негативни или неутрално оптоварени зборови) и сумирање на говорот.

Компонентите за анализа на говор во текст и говор се меѓу врвните карактеристики што голем број клиенти и корисници ги бараа низ годините. **Вредноста на анализирање интервјуа**, **фокус групи**, **говори** или **интеракции снимени на аудио или видео** не може да се потцени. Може да даде нова длабочина и да обезбеди анализа на речиси секоја снимена вербална секвенца.

Функцијата говор во текст (speech-to-text feature) моментално поддржува: англиски (САД, ОК, АУС, Глобал), холандски, француски, германски, италијански, португалски и шпански - со уште повеќе кои ќе следуваат.

Широк опсег на апликации

За академски корисници, говор во текст ќе се покаже како вредна алатка кога се прават интервјуа и собирање податоци од истражување. Истражувачите кои спроведуваат интервјуа или собираат квалитативни податоци можат да препишуваат интервјуа, фокус групи или други снимени разговори користејќи ја функцијата говор во текст.

Способноста за **автоматска транскрипција** и **анализа на говорот** може да го забрза делот за анализа на податоците на повеќето истражувачки проекти, а преку **дискурзивна сентиментација** и **валентна анализа**, може да ги истакне лингвистичките карактеристики кои можеби биле занемарени.



Слика. Функцијата за анализа на говор на iMotions им овозможува на корисниците да ги идентификуваат говорниците и да доделат поединечни секвенци за секој

Вредноста и примената на анализата од говор во текст се протегаат многу подалеку од академските истражувања. Индустриите како што се: **истражувањето на пазарот**, **управувањето со искуството на клиентите** и **анализата на чувствата** сега можат да истражуваат подлабоко во суптилностите на човечката комуникација.

Организациите можат да добијат невиден увид во расположението на клиентите, преку интервјуа, да анализираат големи количини на интеракции во центарот за повици и да извлечат вредни податоци од дискусии во фокус групи или разговори на социјалните медиуми. Способноста за обработка на изговорените податоци им дава овластување на носителите на одлуки да дејствуваат брзо и да прават информирани избори врз основа на сеопфатна лингвистичка анализа.

Како функционира анализа на говор и говор во текст

API на AssemblyAI е платена услуга за обработка базирана на облак. Ова значи дека за да користат анализа на говор во текст и говор во iMotions, корисниците мора да купат **единствен клуч за лиценца за API од AssemblyAI** и да го активираат во софтверот iMotions. Откако тоа ќе се направи, аудио- обработката може да започне.

Користењето на говор во текст и анализа на говор во iMotions е неверојатно едноставно. Едноставно се избера студија што сакаме да ја обработиме и избираме опција говор во текст во менито за постобработка (post-processing), а обработката се одвива во заднина, исто како и сите други функции за постобработка во iMotions.

Откако ќе заврши обработката, резултатите ќе бидат зачувани и видливи како забелешки на нашите податоци. Нашето аудио ќе биде поделено на листа на говорници, секој со низа препишани блокови, како и анализа на чувствата што покажува дали говорникот користи позитивни, негативни или неутрално оптоварени зборови.

На тој начин, покрај практичноста да се расчлени аудио-датотеката на нејзините составни делови и да се транскрибира во процесот, може да се добие и вреден дискурзивен преглед на секој поединечен говорник.

8. Наука за анализа на говор

Гласовна анализа (анализа на говор) е проучување на вокалните карактеристики и обрасци во говорната комуникација. Вклучува испитување на елементи како што се **тон**, **стапка на говор** и **емоционални знаци** за да се добие увид во аспекти како лингвистика, емоции и индивидуални особини. Оваа технологија има примена во области како што се **лингвистиката**, **психологија** и **здравство**.

Да замислиме дека сме во земја во која не сме биле претходно. Ние не го зборуваме јазикот и сигурно не го разбираме. Приоѓате некому, или можеби тој ни приоѓа, и пред да знаеме, тој длабоко нурнува во низа реченици што нема шанса да ги разбереме. Освен што разбираме неколку основни работи. Контекстот секако ќе помогне да се обликува она што е кажано, но има нешто друго што открива многу за тоа што се случува - начинот на кој личноста зборува.



Она што го правиме во тој момент е да го анализираме гласот - можеме да слушнеме промени во тонот, прозодијата и многу повеќе и да заклучиме дали личноста е среќна или тажна, лута или ангажирана. Информациите ни кажуваат многу за внатрешната состојба на личноста. Обично не мора експлицитно да ни се

кажува за емоциите на говорникот, ние автоматски го анализираме гласот и го знаеме одговорот.

Овој феномен, она што денес можеме да го наречеме како **гласовна анализа** , се разбира е познат додека луѓето зборуваат, а првичните обиди за мерење на вокалната продукција траат барем од доцните 1900-ти.

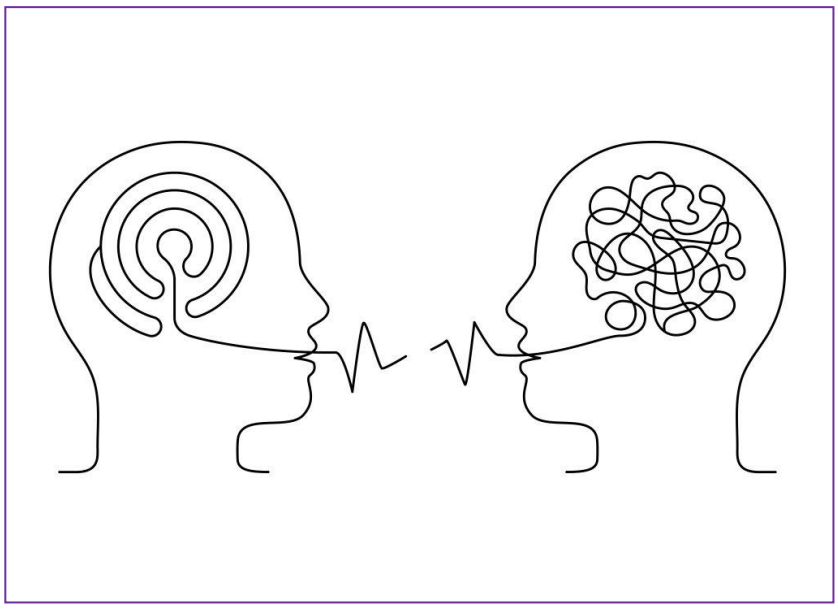
Ова подоцна стана поофицијализирано во 1960-тите со ригорозни научни анализи, но само со доаѓањето на модерните компјутери, гласовната анализа како научно поле навистина се оформи. Сега сме на почетокот на следната промена во науката за гласовна анализа со појавата на методи на вештачка интелигенција.

Но, што е анализа на говор и како може да се користи за подобро разбирање на човечките емоции и однесување? И како помага вештачката интелигенција?

Што е анализа на говор?

Гласовна анализа е процес со кој гласните звуци се мерат и се поврзуваат со дефинирани метрики како што се: **емоции**, **возраст**, **пол** или присуство/отсуство на говор. Клучно, гласовната анализа н**е дава директни информации за употребените зборови**, туку **за начинот на кој тие се произведени**.

Методот вклучува сегментирање на произведениот звук и извлекување на повеќе карактеристики кои потоа може да се проценат независно или заедно. Различни алгоритми работат на различни начини, но обично се собираат и анализираат карактеристики како што се **прозодија**, **стапка** и **интонација**. Тие потоа се користат за да се изгради предвидување на карактеристиките на говорот со повисоки димензии, како на пример дали личноста е **среќна** или **лута**.



Мерења на основна механика на говор и на поширок емоционални резултати се користат за да се обезбедат информации за човечкото однесување. Едно од областите што покажува особено ветување со овој пристап е здравството.

Анализа на говор во здравство

Предноста на анализата на гласот во здравството е двојна. **Прво**, тие се собираат речиси без напор како форма на податоци, кои можат да се снимат (со согласност) уште од првата точка на контакт. **Второ**, тој покажува цврсто ветување како предиктор за широк опсег на невролошки, психијатриски и вокални нарушувања.

Една од првите студии што ја испитуваше врската помеѓу гласот и физиологијата беше спроведена од Meyer Friedman, кој ја популаризираше теоријата на личност од **тип А** и тип Б. Оваа теорија нашироко ги дели луѓето на групи од типот А или типот Б, со тип А “поконкурентен, високо организиран, амбициозен, нетрпелив („competitive, highly organized, ambitious, impatient”) и тип Б како „поопуштен, приемчив и помалку невротичен“ („relaxed, receptive, and less neurotic”).

Истражувачка група откри, преку тест за гласовна анализа, дека може да разликува учесници од тип А и пациентите со срцеви заболувања од учесници од

тип Б. Ова беше земено за да се сугерира дека типовите на личност може да се разликуваат само врз основа на гласот.



Консалтинг, животно осигурување и доктор со црна жена во медицинска канцеларија, чек- листа и резултати. Интервју, здравствена заштита и медицина со пациент и експерт во болница за испит, помош или совет.

Додека теоријата за личности од тип А и тип Б подоцна беше подложена на правилна критика, оставајќи ја суштински разоткриена, пристапот на користење говор со биомаркери за болеста започна. На пример, неодамнешно истражување покажа како биомеханички елементи може да се извлечат само од апликација и да се користат за да се идентификува Паркинсоновата болест.

Истражувањата исто така користеле скенирање на мозокот на пациентите со Паркинсонова болест за да ја покажат врската помеѓу оштетувањето на мозокот во области поврзани со обработка на емоции и емоционален говор, поставувајќи директна поврзаност помеѓу мозокот и гласот во невролошко нарушување.

Истражувањата продолжија да го покажуваат предиктивниот капацитет на говорот (гласот) за идентификување на Паркинсонова болест, Алцхајмерова болест, депресија, амиотрофична латерална склероза (ALS - amyotrophic lateral sclerosis), биполарно растројство, па дури и Ковид болест. Методи за дизајнирање на студија, снимање, екстракција на карактеристики и анализа нашироко вариираат низ студиите, но сепак укажуваат на конзистентна насока - до робусна, рана идентификација на фактори кои влијаат врз здравјето.

Неколку студии и компании го истакнаа ветувањето за автоматска анализа на гласовна состојба (AVCA - automated voice condition analysis) за идни здравствени апликации. Крајно, гласовната анализа покажува големо ветување како ефтин, речиси пасивен начин на рано откривање и дијагноза за широк спектар на болести и нарушувања.

Анализа на говор во интеракција човек - компјутер (HCI)

Гласовната анализа обезбедува уште еден пат за подобрување на здравствената заштита надвор од раната дијагноза – преку помагање на пристапот и емоционалното разбирање на алатките за обезбедување здравствена заштита. Еден таков пример е оној на телездравство, а особено телетерапија.

Истражувањата укажаа на интегрирање на мерки за гласовна анализа во рамки на видео или терапија заснована на повик, обезбедувајќи им на лекарите податоци кои можат да ја водат нивната проценка за пациентот. Една неодамнешна студија покажа како може да се идентификува ризикот од самоубиство – потенцијално давање информации за спасување живот на терапевтот на повик.

Други студии укажаа на начинот на кој интеракциите лекар-пациент можат полесно и објективно да се проценат со методи за гласовна анализа. Ова може да обезбеди податоци за подобрувања засновани на докази што треба да се предложат за идни повици.

Оваа техника на сличен начин може да се примени во реално време на терапевтски чет-ботови (chatbots). Различни написи дискутираа за тоа како овие сè позастапени чет-ботови можат да интегрираат методи за гласовна анализа за

да овозможат одговори кои се поприлагодени на потребите на пациентот преку проценка што оди подалеку од едноставно она што го кажува пациентот.

Оваа емоционално -интелигентна интерактивност што може да се обезбеди во интеграција со гласовна анализа, се разбира, не е ограничена само на терапевтски поставки. Наместо тоа, секоја систематска интеракција која вклучува човечки говор веројатно може да има корист од податоците што можат да ја подобрат емоционалната свест на лицето или вештачката интелигенција што реагира.

Гласовна анализа во UX

Сличен пристап како и кај HCI може да се примени и се применува на UX дизајнот. Ова е конкретно случај со протоколи за гласно размислување, во кои тест-корисникот ги изразува своите мисли додека користи или се движи низ ново дизајниран производ. Производот, кој може да биде што било, од веб-локација, до физички предмет, до пакување, го тестира некој што не го видел претходно. Корисникот едноставно наведува што му паѓа на ум во текот на процесот.

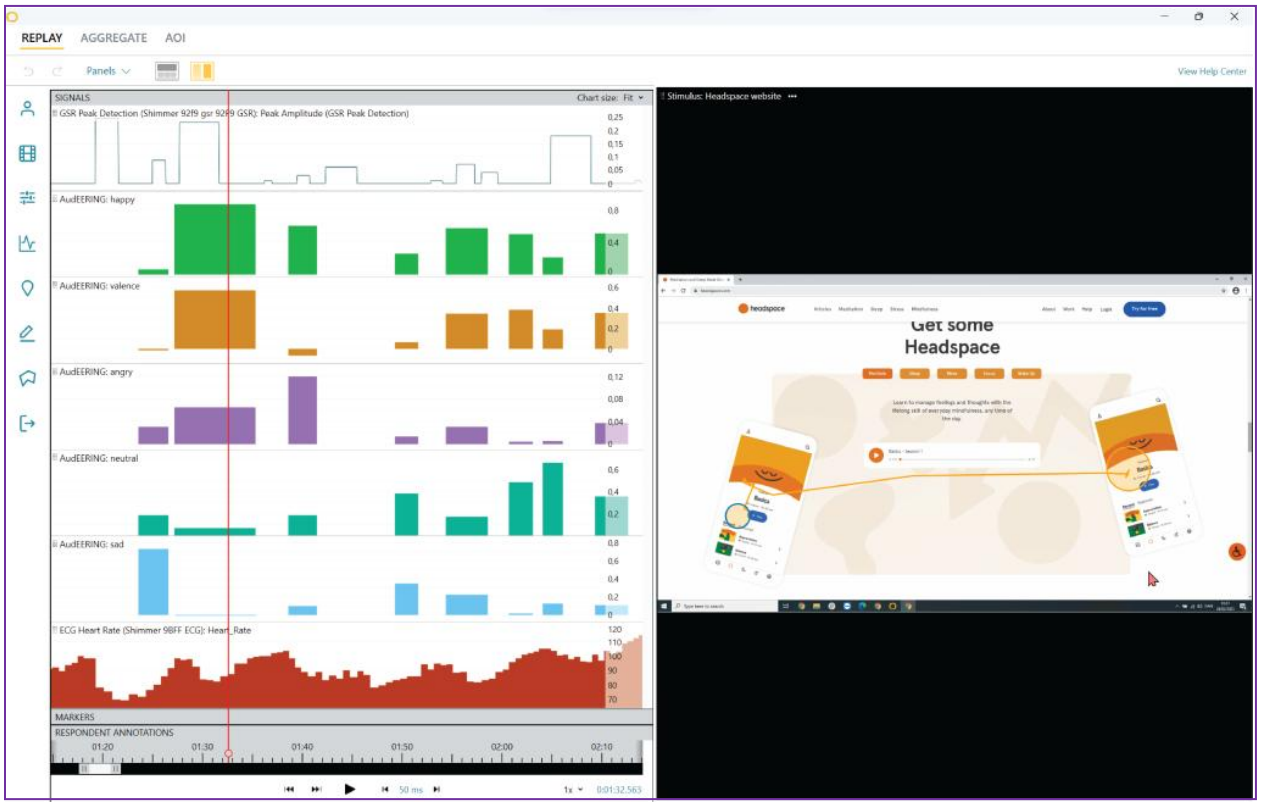
Биосензорите веќе се широко користени за да им овозможат на UX дизајнерите да ги разберат мислите на корисникот надвор од она што тие едноставно го кажуваат гласно. **Следењето на очите** и **анализата на изразот на лицето** може да обезбедат податоци кои обезбедуваат увид **на што всушност луѓето се фокусираат** и **како нивните емоционални изрази се поврзани со тоа**.

Гласовната анализа го претставува следниот логичен чекор во оваа научна прогресија, дозволувајќи им на дизајнерите да ја испрашуваат основната биолошка компонента на протоколите со гласно размислување. Сепак, како што забележаа истражувачите од универзитетот во Торонто и Технолошкиот институт Рочестер, „анализирањето на сесиите на глас често одзема време и труд.“

Истражувачите продолжија да развиваат свој внатрешен метод за машинско учење за анализа на вербализации за време на протокол за гласно размислување. Тука може да се заштеди време дури и со користење на лесно достапни модели како што е уредот од audEERING.

Истражувачите веќе ја комбинираат анализата на изразот на лицето со анализата на гласот за тестирање на „итни ситуации на поморски брод“, “ситуа-

ција од поплава (кризна)“ и „играчи нурнати во игра за виртуелна реалност“, наоѓајќи ветувачки резултати за идни истражувања. Други истражувачи користеле гласовна анализа со протоколи на глас за размислување за испитување на учење во образование базирано на веб и за поставување анализи на искуство во игра.



Откривање на емоции од аудио

Дали некој звучи **среќно**, **луто**, **тажно** или **неутрално**? devAIce® може да разликува четири различни емоции и три димензии на емоции.

Откривање на емоции

Може да се изберат основни модели или приспособени комбинации. Може да се откријат емоции само од аудио, или комбинирани од аудио и текст.

- **Часови за емоции** (од аудио): неутрален, среќен, лут, тажен.
- **Димензии на емоции** (од аудио): активирање, валентност, доминација.

- **Часови за мултимодални емоции** (од текст и аудио): неутрални, среќни, лути, тажни.
- **Димензии на мултимодални емоции** (од текст и аудио): активирање, валентност, доминација.



Слика. Емоции

Voice & Music Activity (Активност со глас и музика)

devAIce® може да прави разлика помеѓу звуци во позадина и човечки глас. Тој е многу ефикасен со мала искористеност на ресурси кога е во мирување. Разликата помеѓу говорот и музиката е достапна како опција.

Карактеристики на говорник

- Возраст
- Пол



Слика. Voice Activity

Откривање на сцена

Разлика до 14 однапред дефинирани акустични сцени:

Транспорт: автомобил, автобус, метро, воз.

Надвор: природа, парк, улица, пешачка улица, голем простор.

Внатре: дом, бања, јавна зграда, ресторан, продавница.

Верификација на говорник

Проценување дали говорникот на снимката е ист како и претходно регистрираниот референтен говорник.

Поддржани се два режима:

- **Режим на запишување (Enrollment mode)**: модел на говорник за референтен говорник се креира или ажурира врз основа на една или повеќе референтни снимки.

- **Режим на верификација (Verification mode)**: претходно креиран модел на говорник се користи за да се процени колку е веројатно истиот говорник да е присутен на дадена снимка.



Слика. Атрибути на говорник



Слика. Акустична сцена

Прозодични карактеристики

- Статистички дескриптори врз основа на висината и гласноста на гласовите.
- Проценка на брзината на зборување во слогови во секунда.

Карактеристики на openSMILE

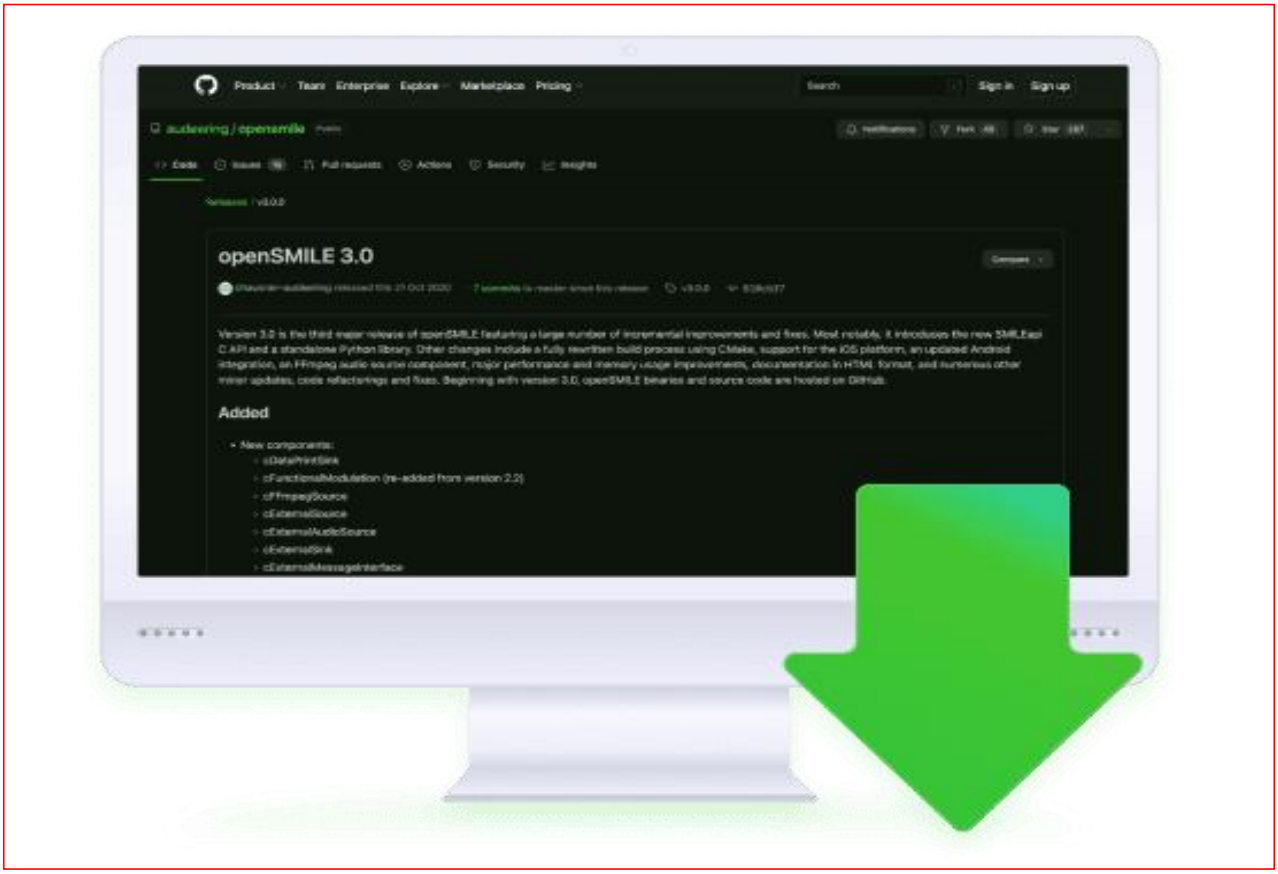
openSMILE е водечка алатка за брза, меѓуплатформска, **open-source** аудио екстракција од auDEERING.



Екстракција на аудио функција со отворен извор (Open-source)

openSMILE (отворена интерпретација на говор и музика со екстракција на голем простор) е комплет алатки со отворен код за екстракција на аудио- функции и класификација на говорни и музички сигнали. **openSMILE** е широко **применет во автоматското препознавање емоции за афективно пресметување**. Open-SMILE е потполно **бесплатен за користење за истражувачки цели.** Особено е насочен кон апликации за говор и музика, на пример, автоматско препознавање на говор, идентификација на говорник, препознавање емоции и откривање акорди.

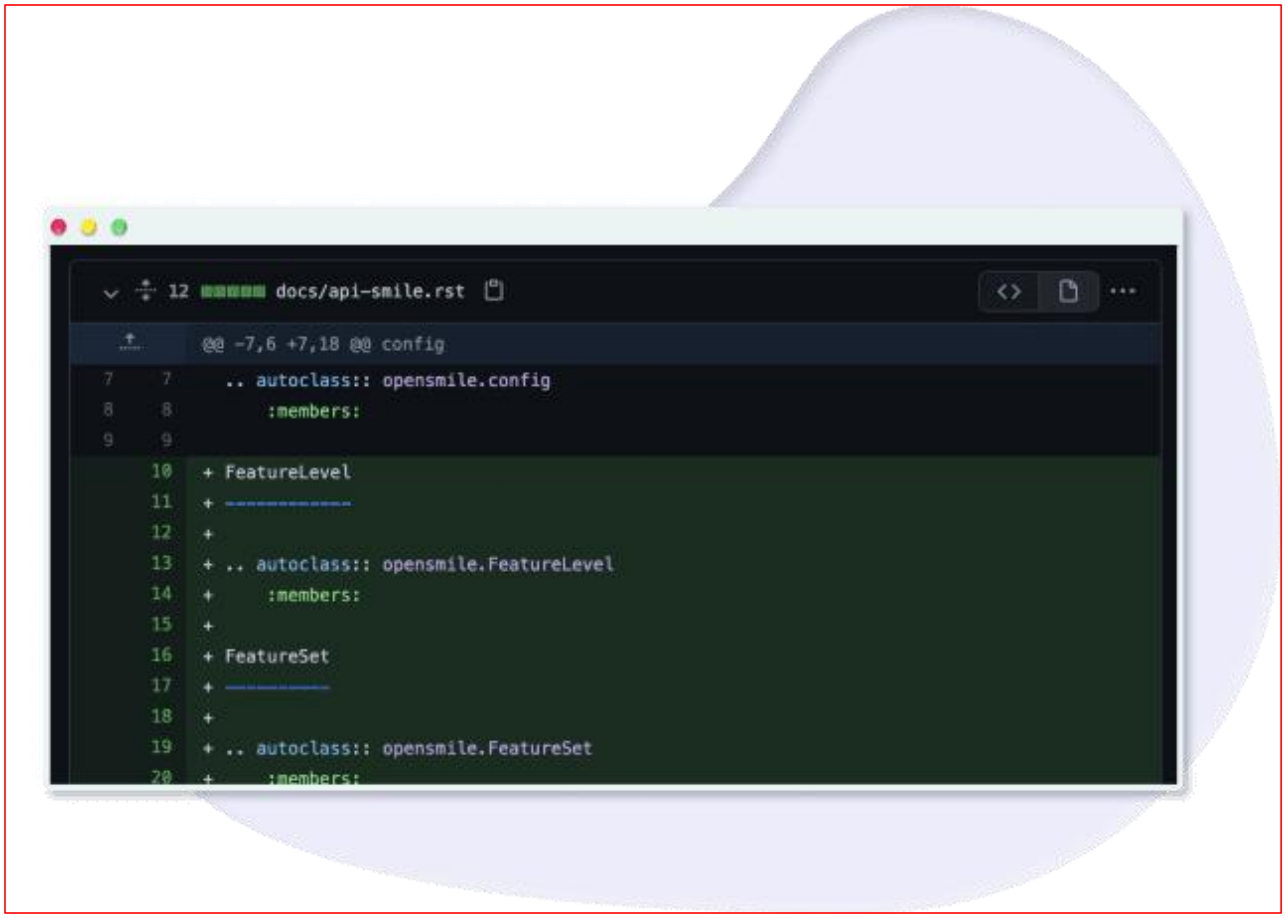
Напишан е чисто во **C++**, има брза, ефикасна и флексибилна архитектура и работи на десктоп, мобилни и вградени платформи како Linux, Windows, macOS, Android, iOS и Raspberry Pi.



openSMILE започна во 2008 година на Техничкиот универзитет во Минхен (TUM), развиен од Флоријан Ејбен, Мартин Волмер и Бјорн Шулер, сите подоцна дел од auDEERING. Во рамките на проектот SEMAINE финансиран од ЕУ, целта беше да се дизајнира виртуелен агент со афективни и социјални вештини. Open SMILE е компонента за анализа на говор и емоции во реално време во овој систем.

Од 2011 до 2013 година, Флоријан Ејбен и Феликс Венингер дополнително го развија OpenSMILE на TUM. Ерик Марчи даде голем придонес за проектот на ЕУ ASC-Inclusion.

Од 2013 година, AUDEERING ги има правата за OpenSMILE и го развива понатаму. Софтверот останува слободно достапен за академска употреба.



Клучни карактеристики на OpenSMILE

Фундаменти (основи)

Ефикасни ресурси: 27k features can be extracted with an RTF of 0.08 Cross-platform (Windows, Linux, Mac, Android, iOS). Брза и ефикасна постепена обработка во реално време. Висока модуларност и reusability of components.

Аудио влез/излез: PCM WAVE датотеки (читање/пишување). Секој формат на медиумска датотека поддржан од FFmpeg (читање). Снимање и репродукција на звук во живо преку PortAudio. Снимање звук во живо преку OpenSL ES/Core Audio на Android/iOS.

Функционирани формати на датотеки: Датотеки со вредност (CSV- Comma-separated value) разделени со запирки (читање/пишување). WEKA ARFF датотеки (читање/пишување). Скриени датотеки со параметри на Markov Toolkit (HTK) (читање/пишување). Формат на датотека со одлика LibSVM (пишува).

Signal processing (обработка на сигнали)

- Windowing Functions (Hamming, Hann, Gauss, Sine, …);
- Fast-Fourier Transform;
- Pre-emphasis filter;
- FIR filterbanks;
- Autocorrelation;
- Cepstrum.

Карактеристики поврзани со говор

- Signal energy;
- Loudness;
- Mel-/Bark-/Octave-spectra;
- MFCC;
- PLP-CC;
- Pitch;
- Voice quality (Jitter, Shimmer);
- Formants;
- LPC;
- Line Spectral Pairs (LSP);
- Spectral Shape descriptors.

Карактеристики поврзани со музика

- Pitch classes (semitone spectrum);
- CHROMA and CENS features;
- Weighted differential.

Обработка на податоци (Data processing)

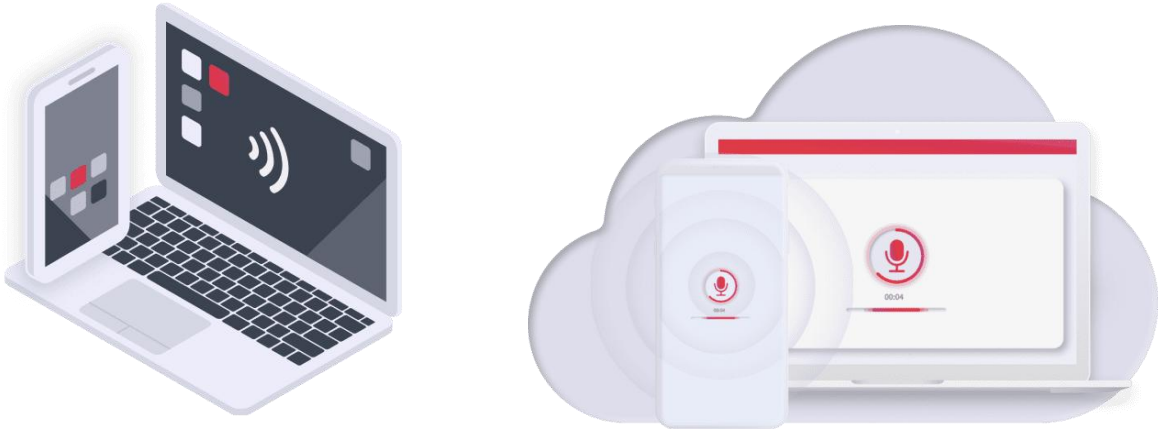
- Нормализација на средната варијанса;
- Нормализација на опсегот;
- Делта-регресивни коефициенти;
- Векторски операции;
- Подвижни просечни филтри (Moving average filters).

Статистички функционалности

- Средни вредности, екстреми;
- Моменти;
- Сегменти;
- Примероци;
- Врвови;
- Линеарна и квадратна регресија;
- Проценти;
- Времетраење;
- Onsets;
- DCT коефициенти;
- Нулта-премин;
- Модулациски спектар.

devAIce® интегрира аудио анализа во софтвер и хардвер. Се користи за откривањето емоции, класификација на животна средина и многу повеќе.

1. **Апликации, програми и веб-апликации**.



Слика. devAIce® Web API: напојуван од облак, природен за веб. **devAIce® Web API** е најлесен начин за интегрирање на аудио вештачката интелигенција во кориснички апликации базирани на веб и облак. Достапни секции за распоредување во просторија за највисоки барања за безбедност на податоци.

2. **Погодни за носење и smart уреди**.



Слика. devAIce® SDK: лесен, природен уред. devAIce® SDK е достапен за сите главни десктопи, мобилни и вградени платформи. Добро функционира и на уреди со ниски пресметковни ресурси, како што се уреди за носење и слушалки

3. **IoT devices**.



4. Аудио вештачка интелигенција за софтвер и хардвер. Софтвер или хардвер - devAIce® е аудио анализа на auDEERING за кој било производ. Може да открие емоции, сцени и многу други функции од аудио. Поддржани се и режими за обработка во реално време и

сериски процеси. Овие модели со вештачка интелигенција работат солидно дури и со ограничена моќност на процесорот.



Подобрување на искуството на нашите клиенти во реално време

Додека возот брза, слушалките Jabra Elite 85h автоматски го активираат поништувањето на шумот. Благодарение на откривањето сцени devAIce®, клиентите на Jabra уживаат во најдобар квалитет на звук без разлика каде слушаат музика.



Слика. Слушалки Jabra Elite 85h се идеално звучно искуство со интелигентна акустична анализа на сцена.

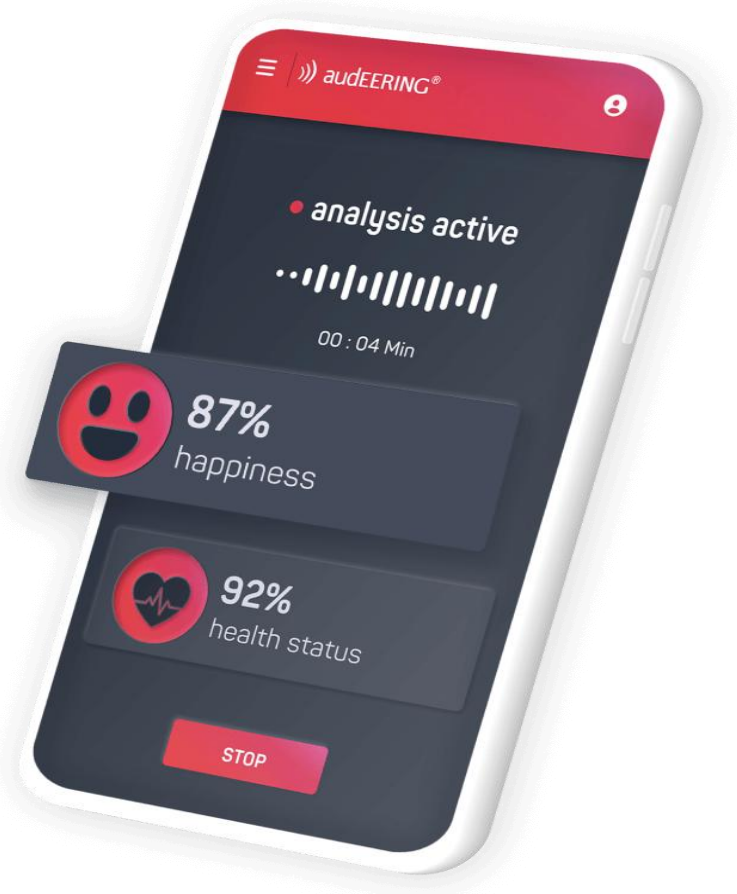
Анализа на говор (глас) базирана на вештачка интелигенција

(AI-based Voice Analytics)

audEERING® е водечки светски иноватор во Voice AI. Оваа технологија може:

- да открие емоции;
- здравствени информации од гласот.

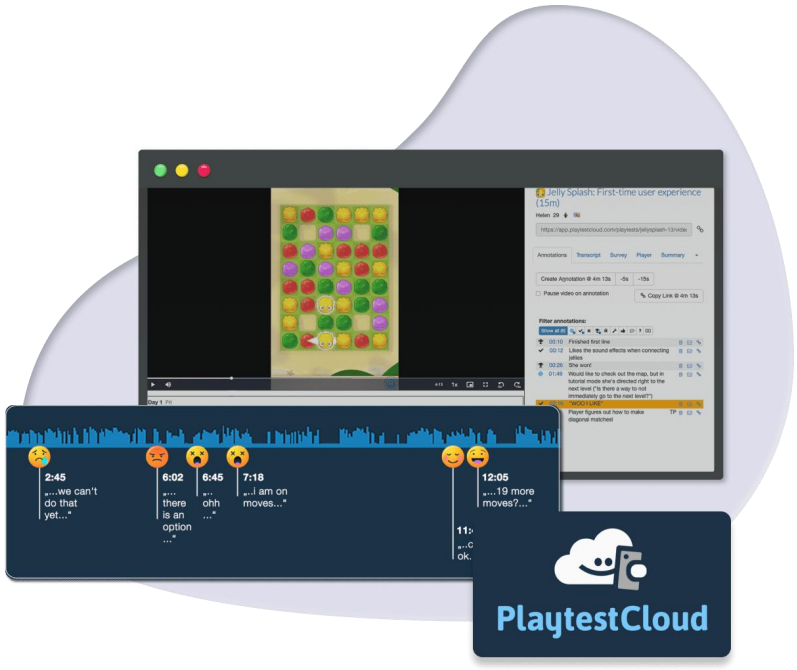
Водечка во светот AI Technology може да следи над 6.000 гласовни параметри во реално време. Истражувачите можат да користат врвни методи на вештачка интелигенција како длабоки невронски мрежи и учење без надзор (unsupervised learning).





Емоционална интеракција со виртуелни ликови

Дали звучиме среќно или досадно? Со вештачката интелигенција од **auDEERING®**, **компјутерите можат да препознаат емоции и да реагираат на нив**. Ова овозможува природни разговори и гласовно контролирани интеракции. Како VR симулатор Iva, кој се користи за обука на персонал во болници.



Слика. Online Testing with PlaytestCloud

Слушање на емоциите на нашите клиенти (Listen to your Customer's Emotions)

Voice AI го носи тестирањето на производот на следно ниво. **Откривањето емоции им дава на клиентите на PlaytestCloud детални информации за нивните тестови за онлајн анализи и истражувања**.

Гласовните биомаркери (Voice Biomarkers) помагаат во медицинскиот третман

Многу психички и физички болести го менуваат гласот. Тие влијаат на респираторниот систем или на мускулите за контролирање на производството на глас. Гласовната вештачка интелигенција открива дури и мали симптоми, што значително ги зголемува шансите за рано откривање. Со **COVID-19** или **Паркинсонова болест** се достигнува точност до **92%**.

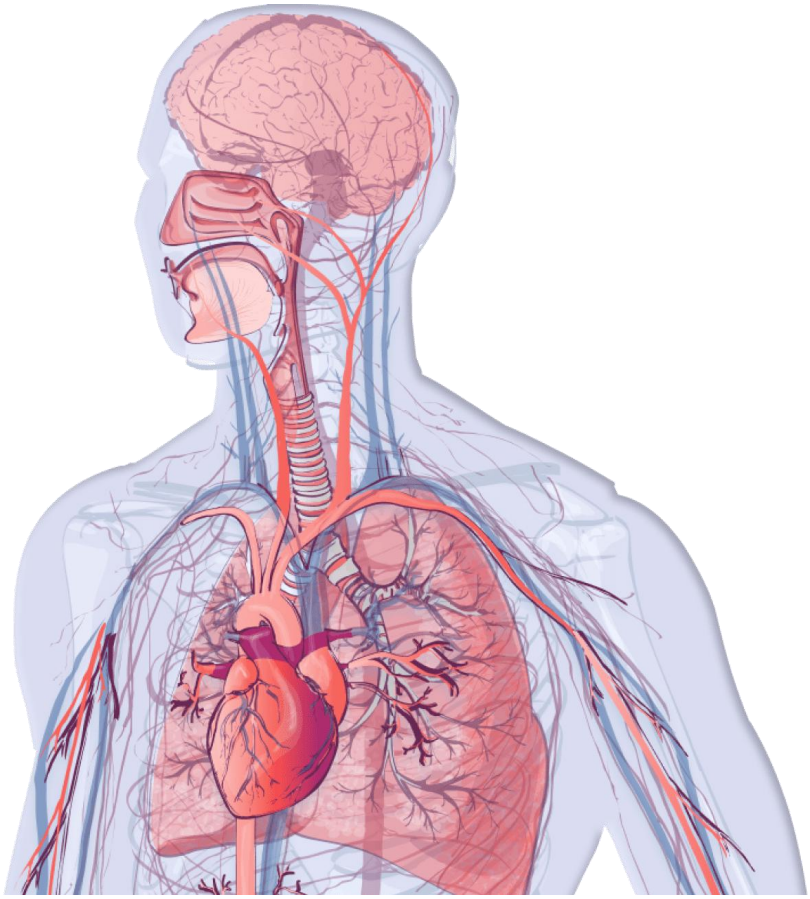


Аудио вештачка интелигенција во здравство: Voice Biomarkers

Откривање болести од гласот? Она што првично звучи како футуристичко сценарио наскоро би можело да стане реалност благодарение на модерната технологија за гласовна вештачка интелигенција (voice AI technology).

Производство на глас и негова поврзаност со болести

Целото наше тело влијае на звукот на нашиот глас. Градите, грлото и главата мора да бидат координирани кога правиме звук. Многу болести влијаат на овој сложен процес. Тие генерираат специфични карактеристики во говорот, кои гласовната анализа на вештачката интелигенција (AI voice analysis) може да ги открие. Овие таканаречени **биомаркери** може да се користат за да се разликува болно од здраво лице.



Чекор 1: **Моторички функционалности**. Во производство на говор се вклучени бројни органи и мускулни групи, на пример: гласните набори, јазикот и муску-

лите на образите. Затоа, моторните нарушувања поврзани со одредена болест се директно забележливи во говорот.

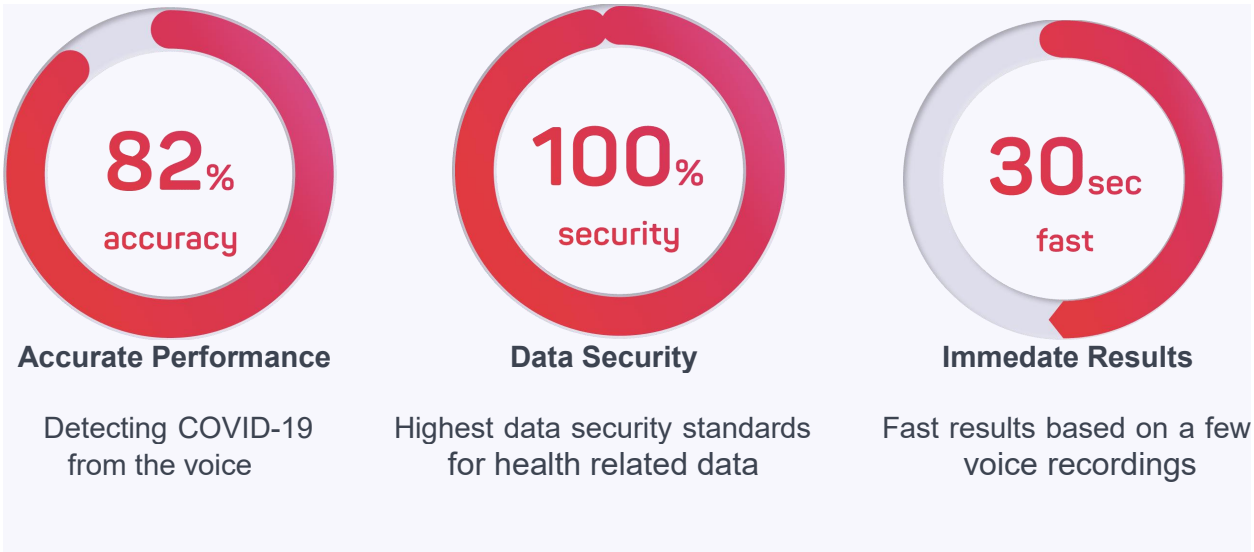
Чекор 2: **Когнитивна контрола и анализа на говорот** (зборувањето). Вербалната артикулација е многу напорна за мозокот. Особено фините моторни мускули мора да бидат совршено контролирани за јасно да се изговараат слогови и зборови. Затоа, когнитивните нарушувања покажуваат карактеристични симптоми рано во текот на говорот.

Чекор 3: **Физиолошки аспекти на гласот**. Звукот на човечкиот глас е директно под влијание на моменталната анатомска и физичка состојба на говорникот. На пример, фреквенцијата на дишење ќе се зголеми додека се занимаваме со спорт.

Откривање COVID-19 со точност до 82%

Респираторните заболувања како СОВИД-19 ги менуваат гласовните карактеристики како што се **интонацијата**, **интензитетот** и **темпото**. Тренирани со звук и гласовни податоци, системите за вештачка интелигенција на audEERING можат да ги препознаат овие промени.

Во соработка со универзитетот во Аугсбург, спроведена е истражувачка студија за тест за COVID-19 базиран на глас. Резултат: Со системот за вештачка интелигенција може да се постигне точност до **82%**.



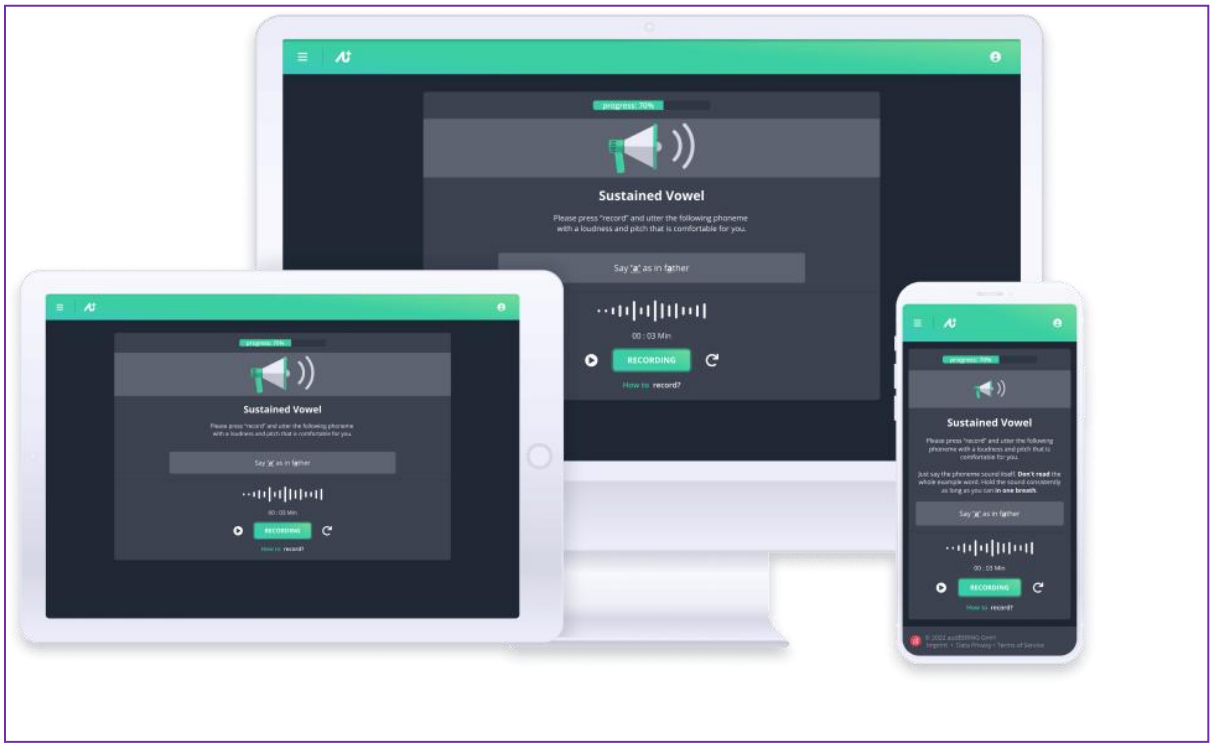
AI SoundLab: Здравствена платформа базирана на веб

Кога се развива здравствена вештачка интелигенција (health AI), прецизноста е најважна. За здравствени апликации, во исто време се бара најдобар квалитет на податоците и највисока можна безбедност на податоците.

Платформа **AI SoundLab** обработува податоци од здрави и болни луѓе според највисоките безбедносни стандарди во Германија. Здравствените податоци збогатени со технологијата на вештачка интелигенција овозможуваат побрза дијагноза од комплетни човечки експерти. **AI SoundLab** може да го следи развојот на здравјето, да обезбедува анализа за дијагноза и да собира податоци за развој на гласовни биомаркери.

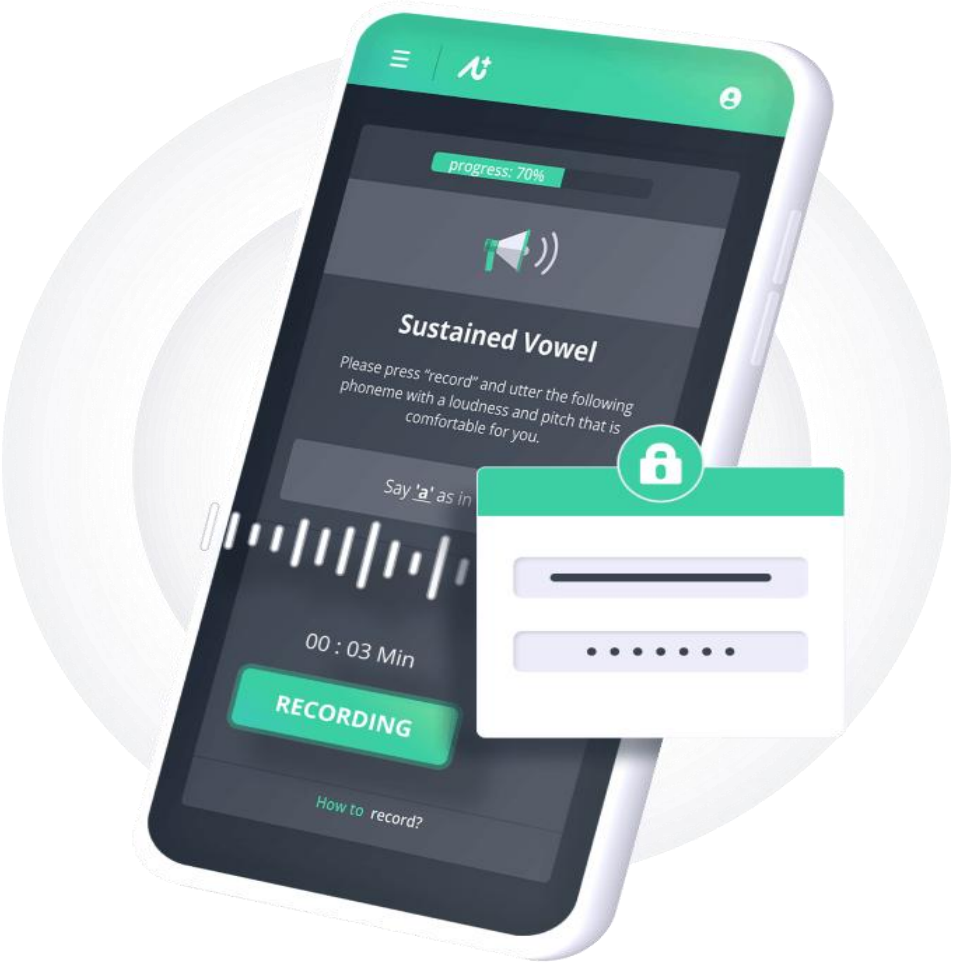
Online платформа AI Soundlab

Прва онлајн- платформа за истражување на говор, здравствени податоци и многу други параметри во една апликација. AI SoundLab комбинира моќна алатка за истражување со најнова гласовна анализа со вештачка интелигенција. Ова овозможува увид од 360° за развој на **гласовни биомаркери**, **автоматска анализа на говор** и многу повеќе.

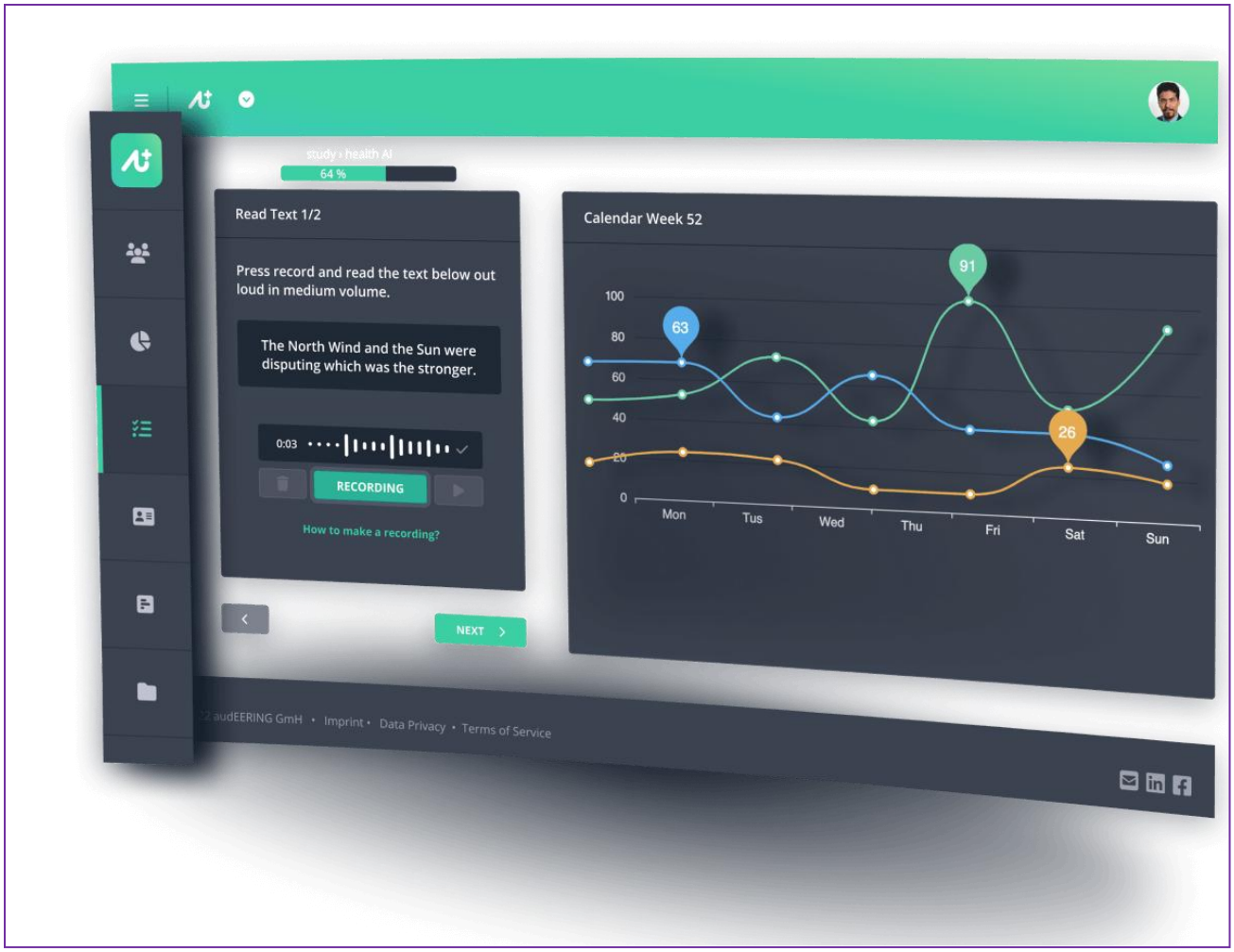


Холистички пристап: Студии за физичко и ментално здравје

AI SoundLab може да го следи развојот на здравјето, да обезбедува анализа за дијагноза и да собира податоци за развој на гласовни биомаркери. Човечки експерти можат да користат здравствени податоци збогатени со технологија со вештачка интелигенција за побрзи дијагнози. **AI SoundLab** ги обработува сите податоци според највисоки безбедносни стандарди во Германија.



AI SoundLab може да открие емоции, атрибути на говорникот, медицински биомаркери и многу повеќе од човечкиот глас.



Слика. Свет во кој voice AI technology може да подобри и спаси животи. Подобрување на благосостојбата на поединците преку обезбедување увид за компаниите во однос на човечките состојби и особини од Voice AI. Подобрување на скринингот, дијагнозата и терапијата на ментални нарушувања преку обезбедување на робусно, прецизно, економично решение Voice AI.

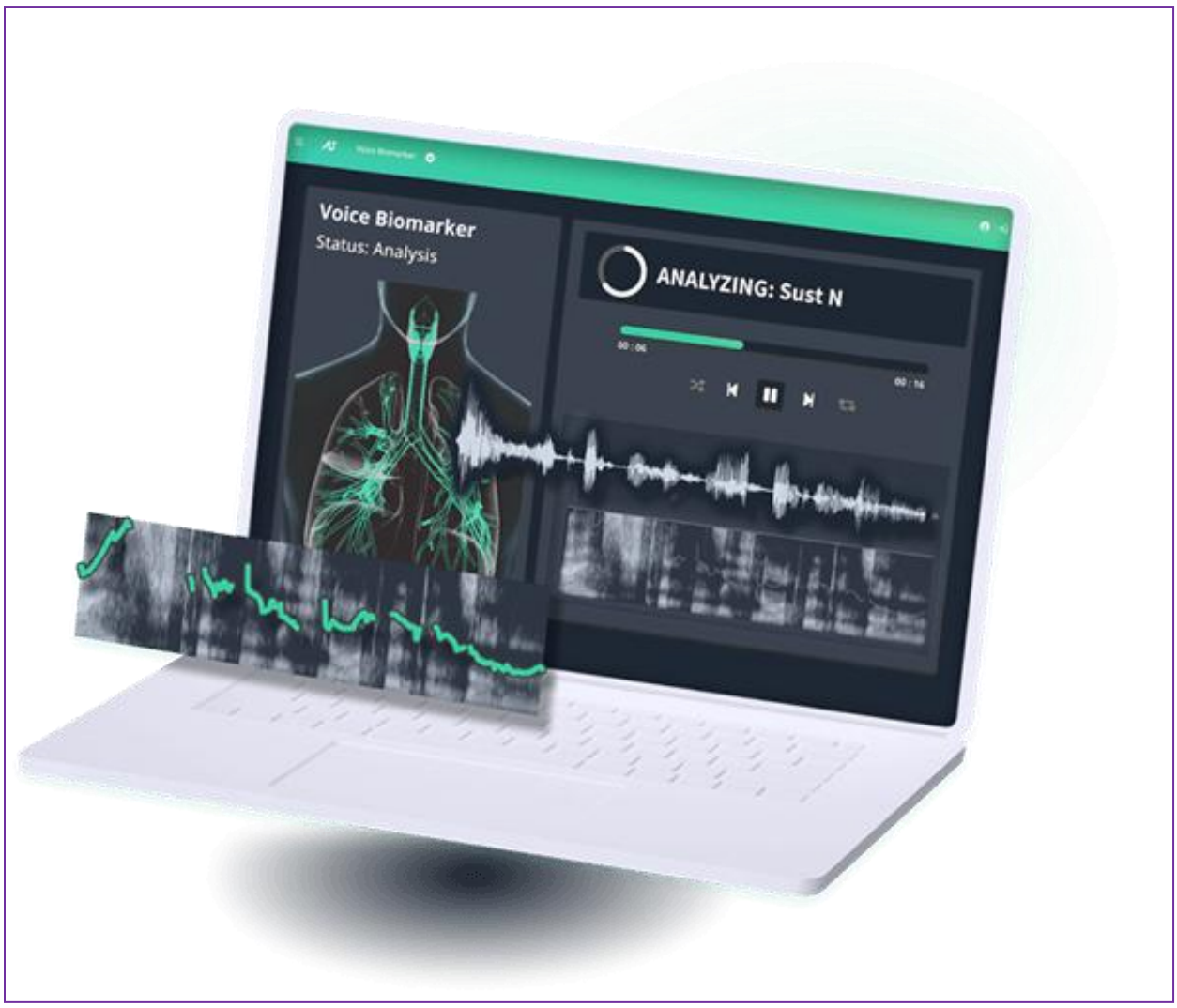
Откривање гласовни биомаркери (Voice Biomarker) и емоции

AI SoundLab е управуван од audEERING®, водечки светски иноватор во гласовна вештачка интелигенција. Платформата **ни овозможува да креираме онлајн анкети поддржани со гласовна анализа**.

Оваа технологија за обработка на глас може да открие различни емоционални состојби и да идентификува гласовни биомаркери поврзани со здравјето.

- Биомаркери поврзани со здравјето добиени од гласот.

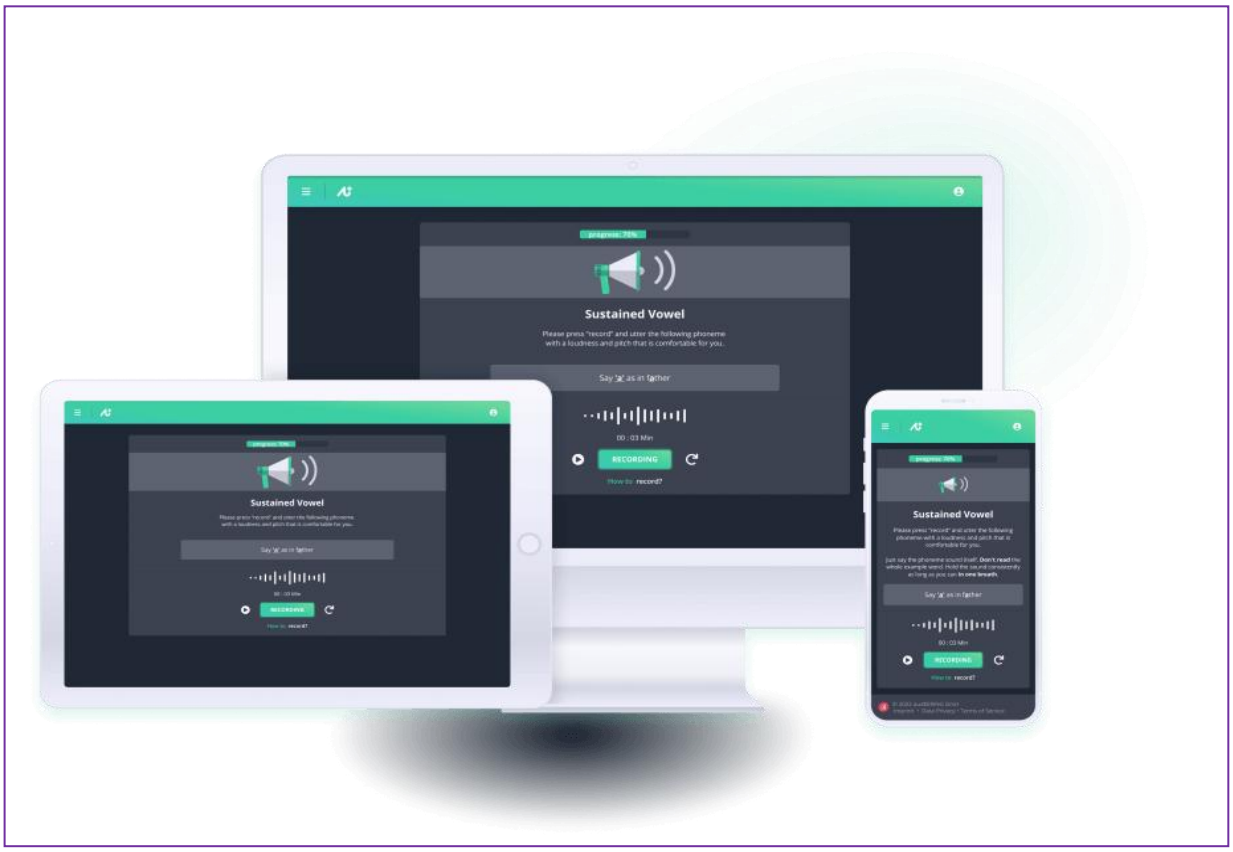
- Идентификување емоции и емоционални состојби.
- Анализа на расположението и следење цело време.
- Поддршка за клинички студии.
- Истражување и евалуација на пазарот.



Создавање своја сопствена гласовна студија со вештачка интелигенција

Собира гласовни податоци за анализа во здравствени или комерцијални гранки. Може да се искористат придобивките од овој систем за обработка на глас

за да се анализираат добиени податоци. Може да се користат и прилагодат постојни различни шаблони за прашања за да одговараат на поставени цели.

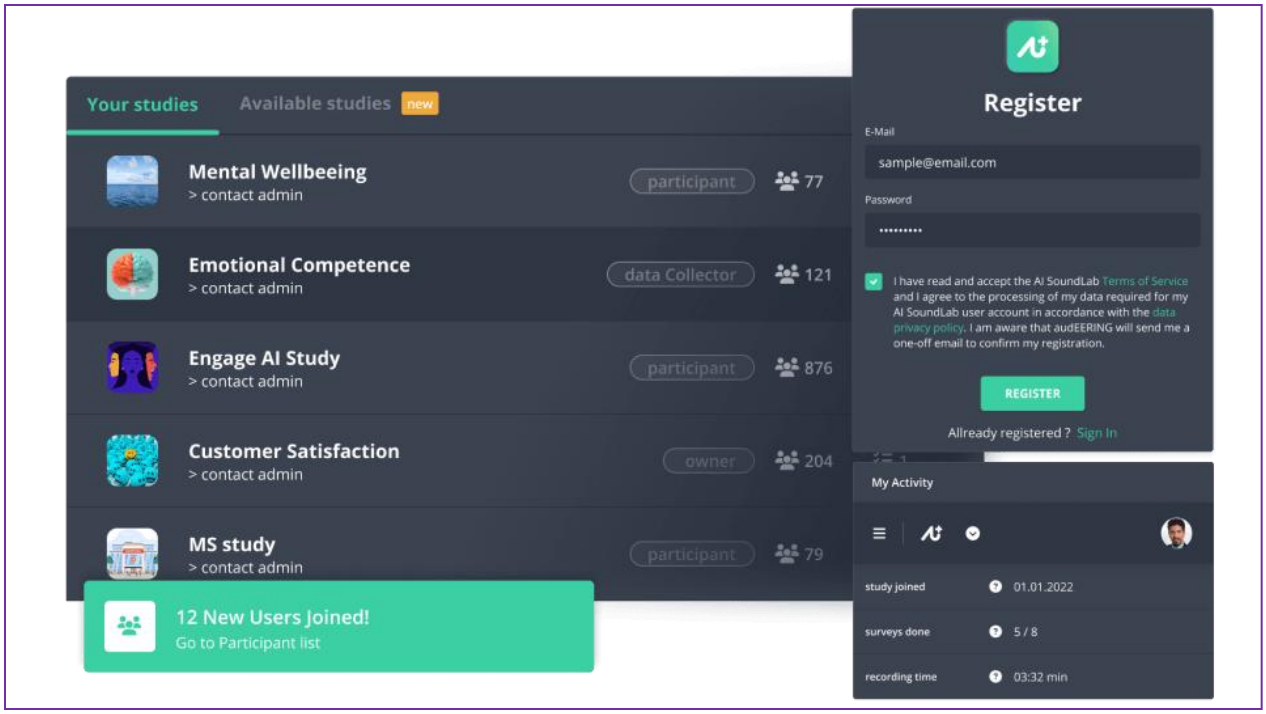


Лесна регистрација и менаџирање со испитаници во истражување

Може рачно да се додаваат испитаници или да се користи саморегистрација преку е-пошта. Процесот на регистрација со двојно пријавување им овозможува на испитаниците самостојно да прават анкети: во секое време, на кое било место.

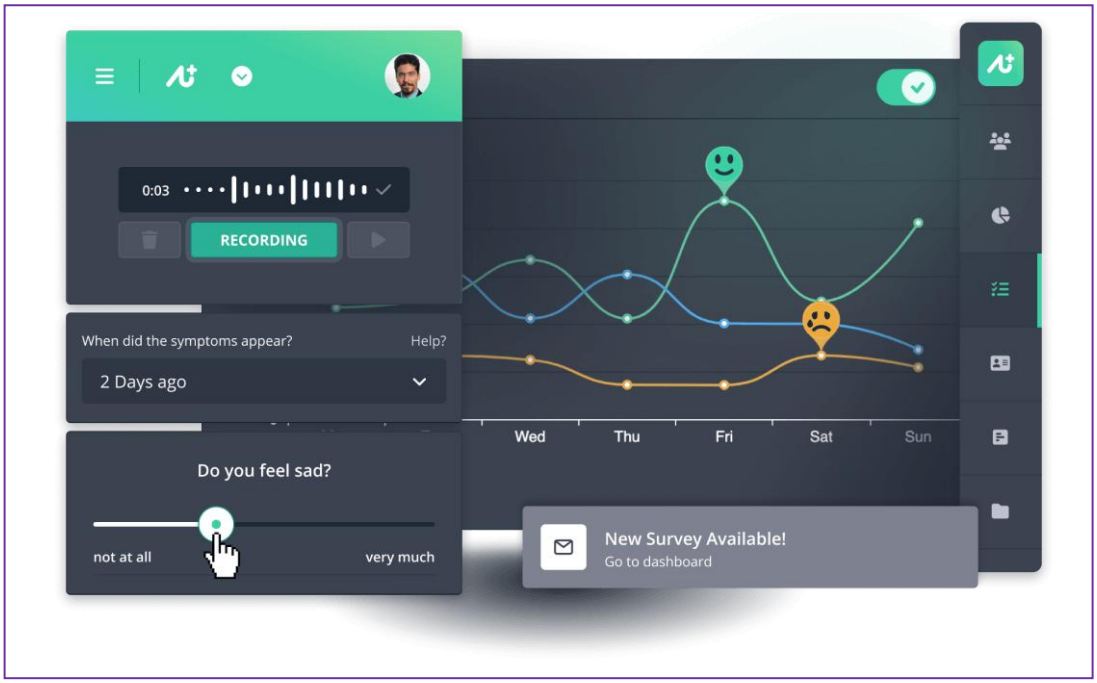
Супервизорите на студијата можат лесно да управуваат со своите испитаници и да го следат целокупниот напредок на студијата.

- Поканување на испитаници преку е-пошта или со код.
- Овозможување учество на анонимни корисници.
- Поединечни анкети, повторувачки или долгогодишни студии.



Обработка на резултати во реално време

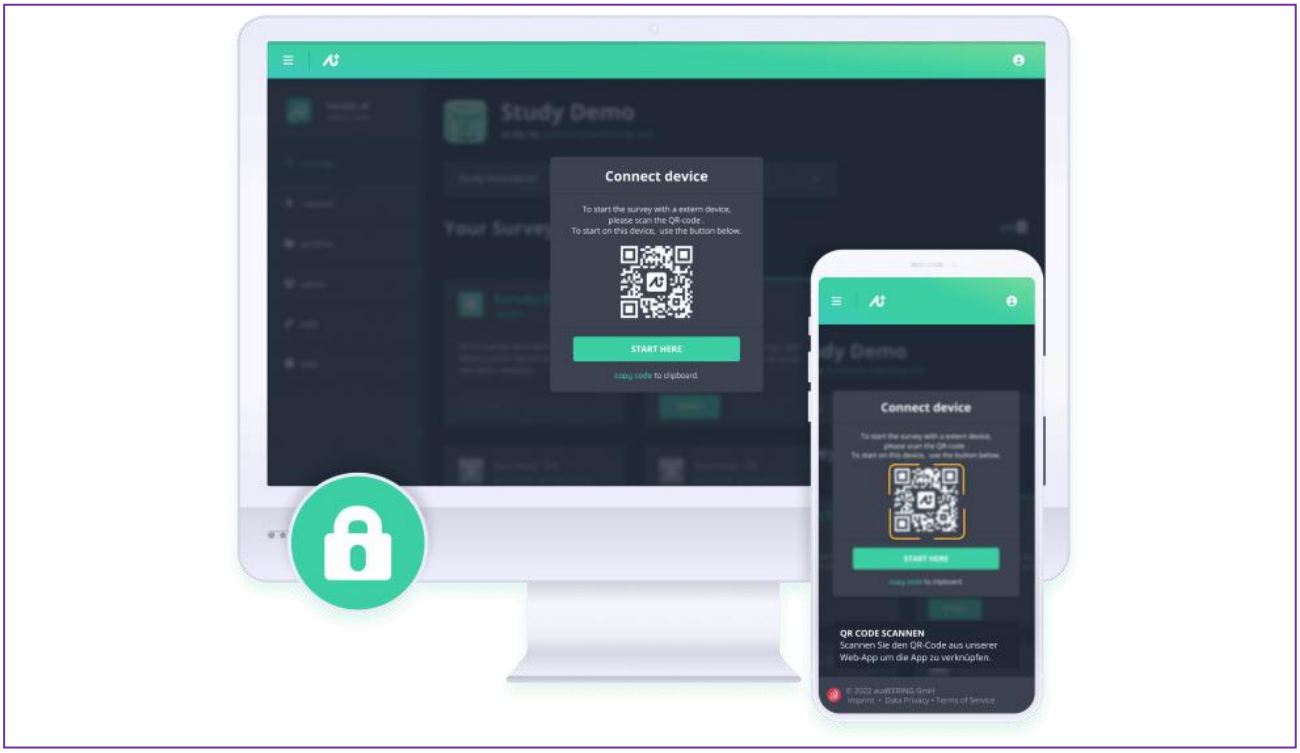
Модулот за обработка ги анализира дојдовните податоци во реално време. Можат да се гледаат поединечни резултати за секој испитаник или, збирна статистика (од сите испитаници) на целата студија.



Највисока безбедност на податоци за десктоп и мобилна употреба

Гласовните и здравствените податоци се чувствителни информации и треба да се придржуваме до највисоките стандарди за да ги заштитиме, без разлика дали со лаптоп во безбедна мрежа или во движење со паметен телефон.

- Во согласност со GDPR.
- Целосна контрола врз нашите податоци.
- Отповикување согласност за обработка во секое време.
- Обработка на сервери во Германија.



Слика. AI SoundLab нуди ефикасно решение за собирање податоци, преглед и последователна анализа

Апликација за здравје базирана на глас

Креативна визија со AI SoundLab е значително да го подобриме индивидуалното здравје. Оваа технологија ќе направи голема разлика за зголемените

невродегенеративни и неврокогнитивни болести. Гласовните биомаркери базирани на вештачка интелигенција ќе овозможат:

- скрининг на болеста;
- рано откривање на симптомите;
- засилена терапија во иднина.

Откривање на болести со анализа на глас со 92% точност

И физичките и психичките болести се рефлектираат директно во гласот. audEERING развива прецизни гласовни биомаркери кои може да се користат за дијагностицирање на невронски болести како Паркинсонова болест или вирусни инфекции како СОВИД-19.



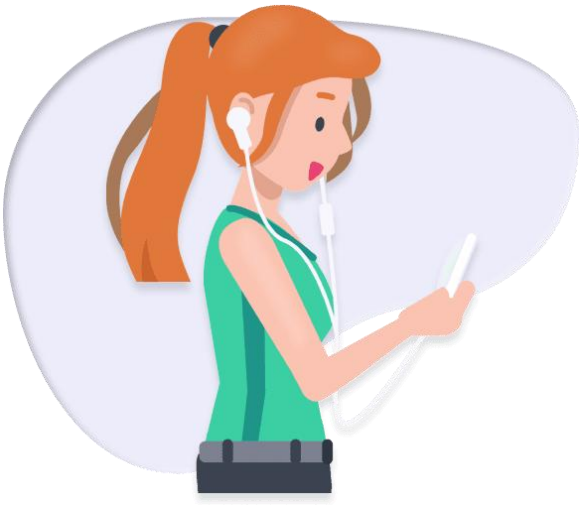
Базиран на глас

Брз тест на паметен телефон

Со гласовни биомаркери, би било можно да се открие COVID-19 преку глас за неколку минути. Користење само еден сензор: микрофон.

Чекор 1: Собирање податоци

За да изградиме гласовен биомаркер за COVID-19, потребни ни се здравствени податоци за да ги нахраниме нашите модели со вештачка интелигенција. Затоа, е создадена студија за COVID-19 за AI SoundLab.



Чекор 2: Тренирање на модел со вештачка интелигенција

Овој прототип веќе има точност од 82%. За да се подобри ова, истражувачи за вештачка интелигенција ги користат собраните податоци. Тие го обучуваат моделот на вештачка интелигенција со милиони точки на податоци и го подобруваат алгоритмот.

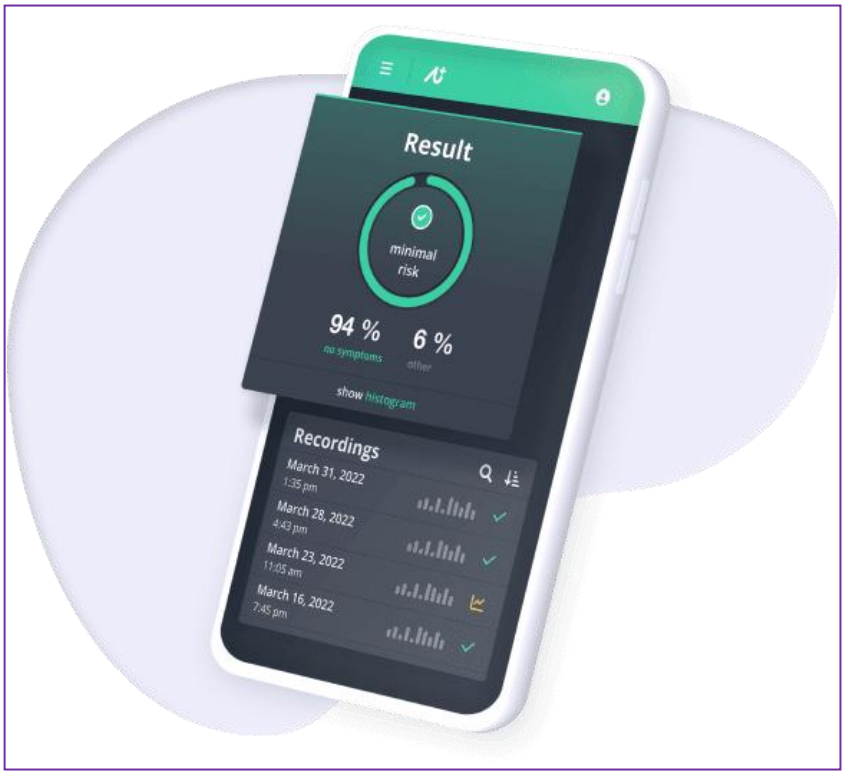
Целта е да се создаде биомаркер за глас за откривање на COVID-19 со многу висока точност.



Чекор 3: Креирање (создавање) апликација

Гласовните биомаркери не само што се многу точни, туку и моделите со вештачка интелигенција кои стојат зад нив се исто така многу лесни. Тие можат да се извршуваат на процесори на мобилни уреди како и на паметни телефони.

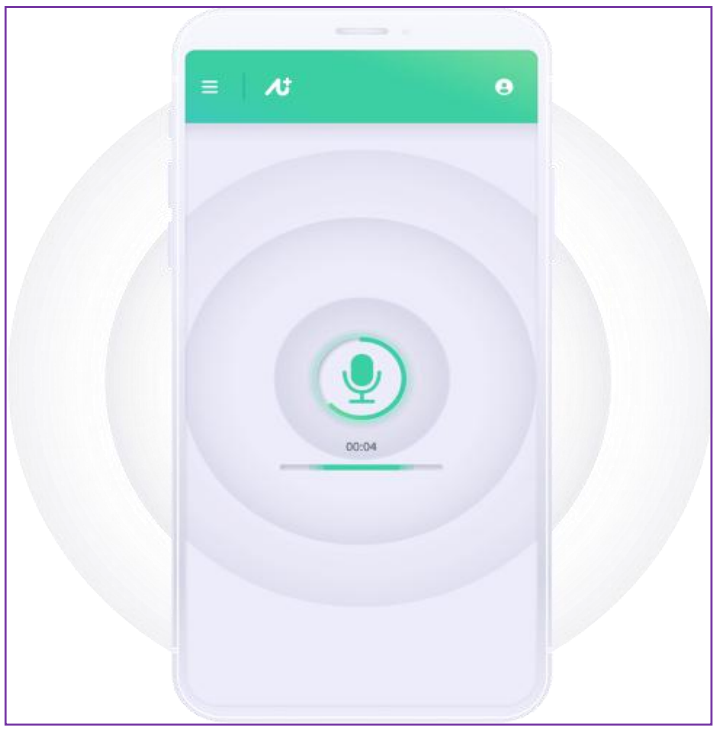
За таа цел, audEERING создаде мобилна апликација, која собира аудио податоци и дава резултати.



Апликација со аудио вештачка интелигенција против COVID-19

Паметните телефони овозможуваат снимање со доволен квалитет и имаат доволно процесорска моќ да ги пресметаат резултатите во најкус можен рок.

Без неуредни брисеви, без пластичен отпад, како и без тестови за еднократна употреба. Достапно и скалабилно решение, дури и за земји со пониски буџети за здравствена заштита.



ПРАКТИЧНО ИСТРАЖУВАЊЕ - КОРИСТЕЊЕ БИОСЕНЗОРСКА ТЕХНОЛОГИЈА ЗА СЛЕДЕЊЕ НА ОЧИ

Истражувањето за мојот магистерски труд го реализирав во современа лабораторија за биосензорски истражувања - невромаркетинг и невропсиологија, на Меѓународен Славјански Универзитет во Свети Николе, во период од 07.05.2024 до 23.05.2024 година, под менторство на проф. д-р Сашо Манасов. Во истражувањето користев биосензорски систем за следење на очи - хардвер и софтвер: мерење на активноста на очи на испитаници (студенти) при следење на одредени визуелни стимули кои се прикажани на компјутерски екран, за да се процени визуелното внимание (на кои елементи од стимулот испитаниците обрнуваат најголемо внимание).

Следењето на очите на испитаници овозможува да се измери нивното визуелното внимание, метрика која објективно следи **КАДЕ**, **КОГА** и **ШТО** гледаат тие. Во моето истражувањето се анализираат и прикажуваат следните метрики добиени од биосензорска технологија за следење на очи:

- **Топлински мапи** (**heat maps**): откриваат дистрибуција на визуелно внимание и претставуваат практичен метод да се визуелизира кои елементи на стимулот привлекуваат одредено ниво на внимание.
- **Области на интерес** (**Areas of interest – AOI**): претставуваат кориснички однапред дефинирани бафер зони (подрегиони) на презентиран стимул: се користат за проценка на перформанси на две или повеќе специфични области на ист видео стимул и споредувања на одредена таргет група на испитаници. Исто така го пресметува количеството на време кое испитаниците го поминале гледајќи одреден сегмент од интерес (dwell time – time spent). Поголемо визуелно внимание во одредена бафер- зона на стимулот јасно укажува на **повисоко ниво на интерес** (**привлечност**).
- **Низи на фиксации (fixation sequences)**: позиција на фиксација каде гледаат испитаниците и информации за време и колку време испитаниците поминуваат во гледање определен елемент од стимулот. Даваат увид во редослед на визуелно внимание каде испитаниците погледнале прв, втор, трет пат, итн.
- **Дилатација (проширување и собирање на зеница).** Дилатација на зеница е пропорционална на **ЈАЧИНАТА НА ВОЗБУДАТА – покажува ниво на емоционална возбуда**.
- **Фреквенција на трепкање:** Ниска фреквенција е поврзана со повисоко ниво на концентрација, а висока фреквенција на трепкање е индикативна за поспаност, пониско ниво на фокус и пониско ниво на концентрација.

Во ова истражување користев два видео стимули:

- **Стимул 1**: Избор на чоколадо (од четири понудени);
- **Стимул 2**: Избор на чипс (од три понудени).

Преку ова истражување сакам да истражам каква храна најмногу сакаат да конзумираат студентите. Притоа користев метод на истражување:

Биосензорско истражување (потсвесно) со користење биосензорски систем за следење на очи и со анкетирање на испитаници (со заокружување покажуваат кој производ најмногу им се допаѓа).

Анализа на резултатите од биометриско следење на очи:
поединечни и збирни анализи

а) Топлински мапи (Heat maps) – Attention maps: **Поединечни** анализи

Топлинските мапи прикажуваат збирни анализи (на сите испитаници заедно) на визуелно внимание, а можат да прикажат и поединечни анализи (на секој испитаник одделно).

- **Црвени** области: Покажуваат голем број на точки на поглед – зголемено ниво на интерес.
- **Жолти и зелени** области: Покажуваат помалку точки на погледот – помало ниво на интерес.

Области кои не се обоени, се оние кои не биле посетени од погледи на испитаници.

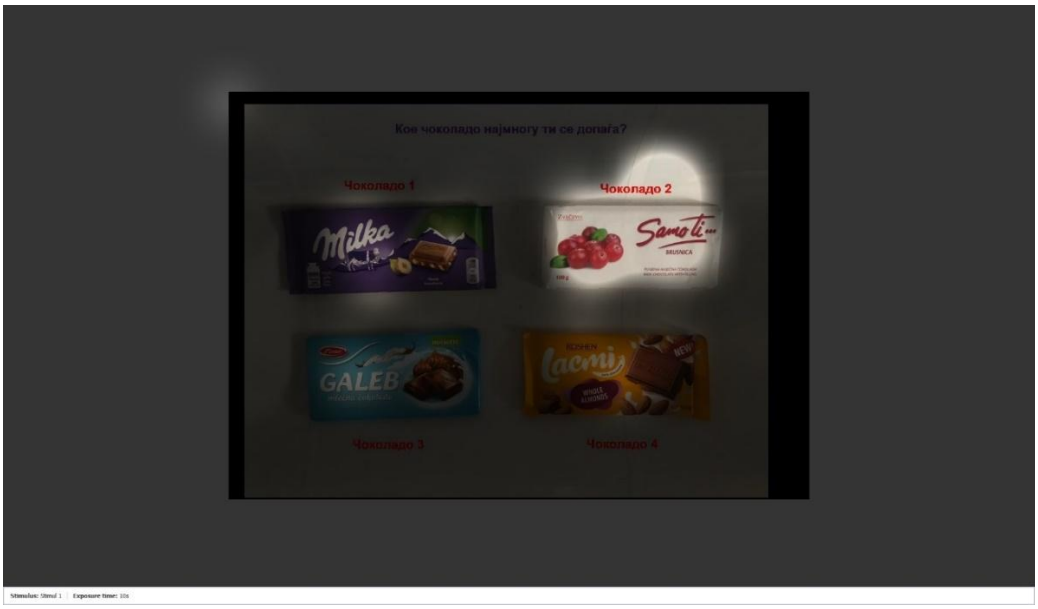
На слика 1 е даден визуелен приказ на Стимул 1 при негово следење од испитаник 9. Се забележува дека овој испитаник најголем интерес покажа за **Чоколадо 2**, и значајно помал интерес за останатите понудени производи. Ова покажува дека овој испитаник нема никакви дилеми кое чоколадо би го избрал од понудените. На анкетниот прашалник овој испитаник го заокружил исто така Чоколадо 2, што покажува дека неговото рационално и потсвесно размислување не се разликуваааат. Ова практично покажува дека неговиот одговор ќе биде ист во следните два случаи:

- Треба брзо да одговори кој производ најмногу му се допаѓа: ова е потсвесно размислување кое се определува со биосензорска технологија.
- Дава одговор по одредено време на размислување (рационално).

На слика 2 е прикажана Топлинска мапа (spotlight) за иста опсервација како и слика 1. Се забележува дека фокусираното внимание на испитаникот е пропорционално со нивото на осветленост. Овде се забележува дека најголема осветленост е на Чоколадо 2



Слика 1. Топлинска мапа на Стимул 1, испитаник 9



Слика 2. Топлинска мапа (spotlight) на Стимул 1, испитаник 9

На слика 3 е прикажана топлинска мапа при следење на екрански стимул 2 на испитаник 5. Овој испитаник најголем интерес покажа за вториот производ (Chipsy 2), значително помал интерес за третиот производ, а најмал интерес за првиот производ. Црвената боја генерирана на топлинската мапа е првичен и најважен индикатор каде испитаникот најдолго го задржал својот поглед – производ што најмногу му се допаѓа.

На слика 4 е прикажана Топлинска мапа (shadow) за испитаник 5, стимул 2. Во оваа топлинска мапа индикатор за количество на фокусирано внимание е површината на генерирана сенка (област каде испитаникот го задржал погледот при следење на стимулот – траење 10 секунди). Се забележува дека најголема површина на генерирана сенка има на вториот производ (Chipsy 2), помала сенка на третиот производ, а нема сенка на првиот производ (најмал интерес од овој испитаник за овој производ).

При пополнување на Анкетен прашалник, овој испитаник го заокружил вториот производ, што јасно покажува дека не постои дискрепанца помеѓу неговото рационално и потсвесно размислување



Слика 3. Топлинска мапа, стимул 2, испитаник 5



Слика 4. Топлинска мапа (shadow), стимул 2, испитаник 5

На слика 5 е прикажана топлинска мапа на испитаник 11 за стимул 2. Се забележува дека овој испитаник најголем интерес покажа за производ 1 и производ 2, и нешто помал интерес за производ 3. Ова покажува дека тој нема производ кој го издвојува (најмногу му се допаѓа). За производот 1 покажал најголем интерес (38.8% од времето го следел), а за производ 2 (29.2% од времето). Овие блиски вредности сугерираат дека неговиот избор е помеѓу овие два производи.

На Анкетниот прашалник овој испитаник го заокружил производот 2 (Chipsy 2). Ова покажува дека постои дискрепанца (1-2) помеѓу овие два производи. Ова значи дека при рационално размислување (анкетен прашалник – има подолго време за размислување) испитаникот се одлучил за Производ 2. Додека, при потсвесно (брзо, биосензорско) размислување (одлучување), тој дава незначителна предност на производ 1 (Chipsy 1) (Приказ: Табела за Споредување на поединечни вредности – на крајот на магистерскиот труд).

На слика 6 е прикажана топлинска мапа каде површината на сенката е пропорционална на нивото на интерес на испитаникот.

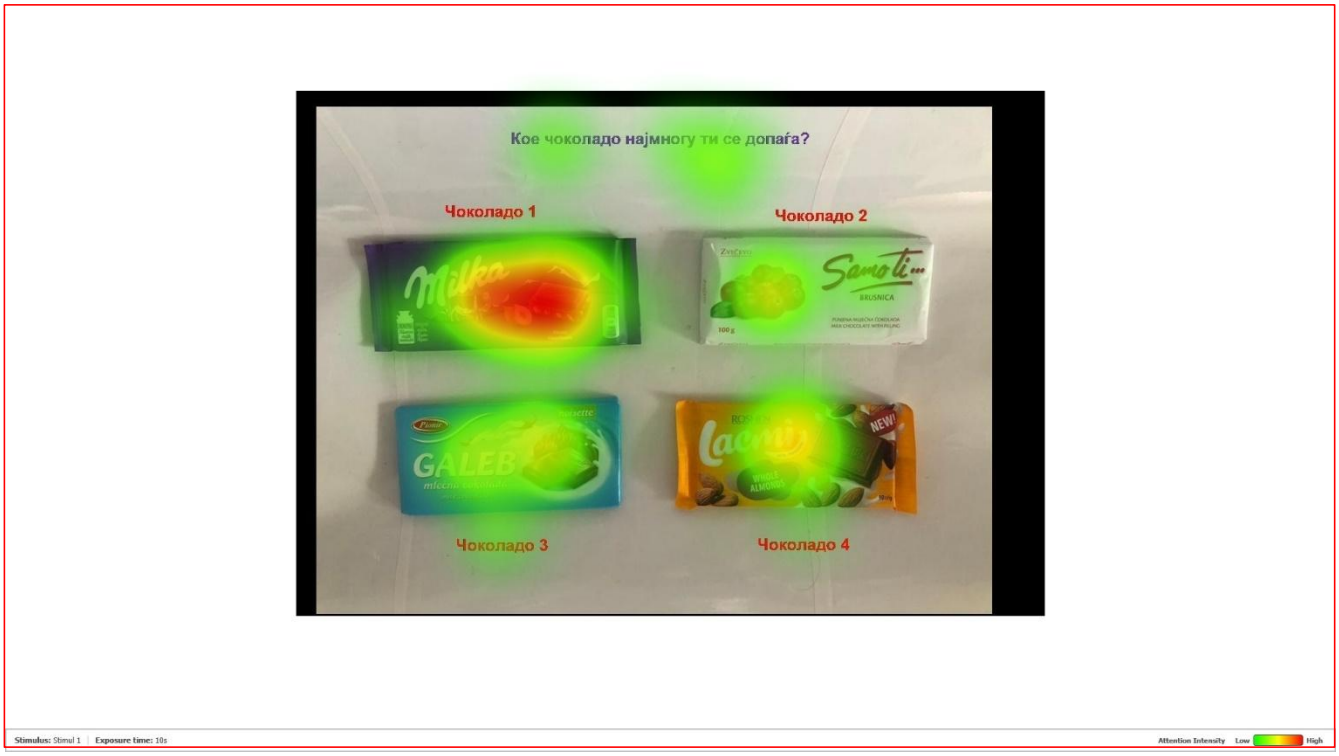


Слика 5. Топлинска мапа, стимул 2, испитаник 11



Слика 6. Топлинска мапа (shadow), стимул 2, испитаник 11

На слика 7 е прикажана топлинска мапа на испитаник 1 кој следел екрански стимул 1. Може да се забележи дека овој испитаник најголем интерес покажа за првиот производ (Чоколадо 1), многу помал интерес за чоколадо 4, а најмал интерес за останатите производи. Овој испитаник има свој омилен производ (од понудените четири), што го потврдил и на анкетниот прашалник (го заокружи Чоколадо 1). Двата метода на истражување не покажуваат дискрепанца при опсервација на овој стимул.



Слика 7. Топлинска мапа, стимул 1, испитаник 1

На слика 8 е прикажана топлинска мапа (за ист испитаник и ист стимул како на слика 7) каде фокусираното внимание (ниво на интерес) е претставено преку прикажана сенка. И овде се забележува дека сенката има најголема површина на Чоколадо 1, а значајно помали површини на сенки на останатите производи. Овој начин на претставување е втор (очигледен) визуелен приказ за дистрибуција на фокусирано внимание.



Слика 8. Топлинска мапа (shadow), стимул 1, испитаник 1

б)Топлински мапи (Heat maps): Збирни анализи

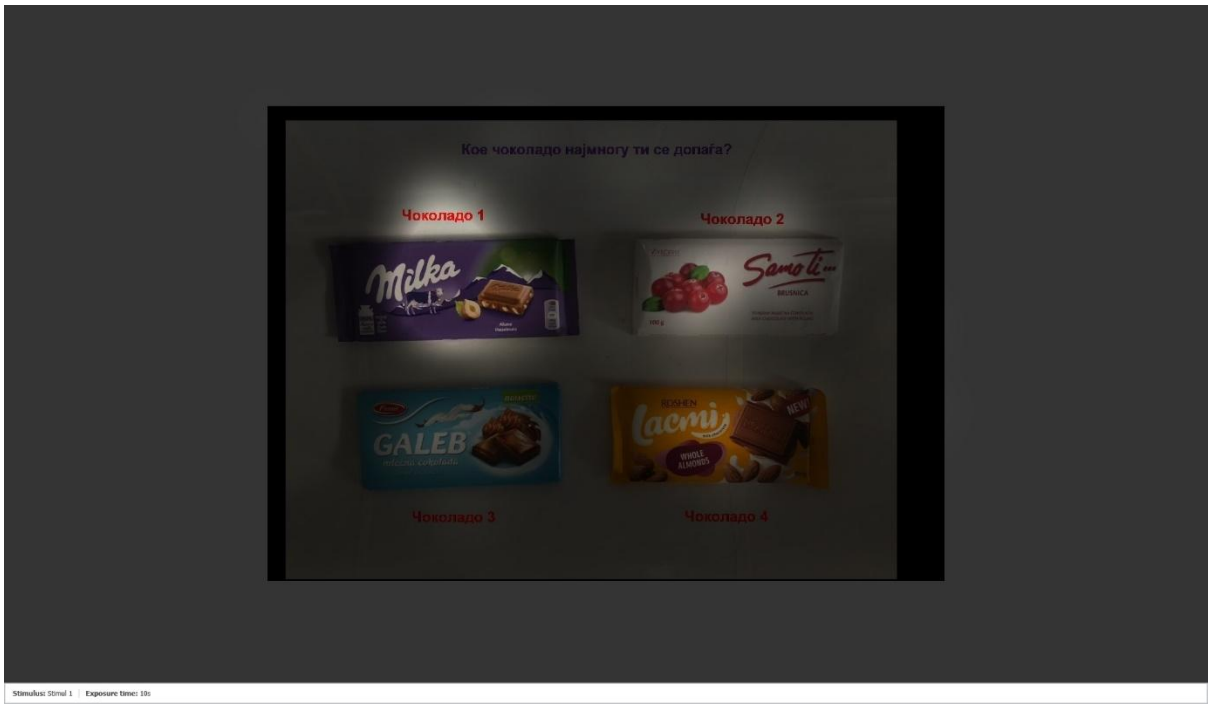
Збирна топлинска мапа е биосензорска метрика која најмногу се користи во невромаркетинг истражувања. Таа покажува збирни резултати од сите испитаници заедно и претставува релевантен параметар за она што најмногу е од интерес за одредена целна група.

На слика 9 е прикажана збирна топлинска мапа од сите испитаници заедно за стимул 1. Може да се забележи дека најголем интерес покажува Чоколадо 1. На анкетен прашалник, 10 од 11 испитаници го заокружиле чоколадо 1 – тоа изнесува 90.9%. Кај биосензорско мерење со користење на биосензор за следење на очи, 9 од 11 испитаници го избрале чоколадо 1 што претставува 81.8%. Ова се истражувачки резултати кои покажуваат каков производ преферира оваа група на испитаници.

На слика 10 е прикажана збирна топлинска мапа (spotlight style) за стимул 1 каде нивото на осветленост е пропорционално со ниво на фокусирано внимание.

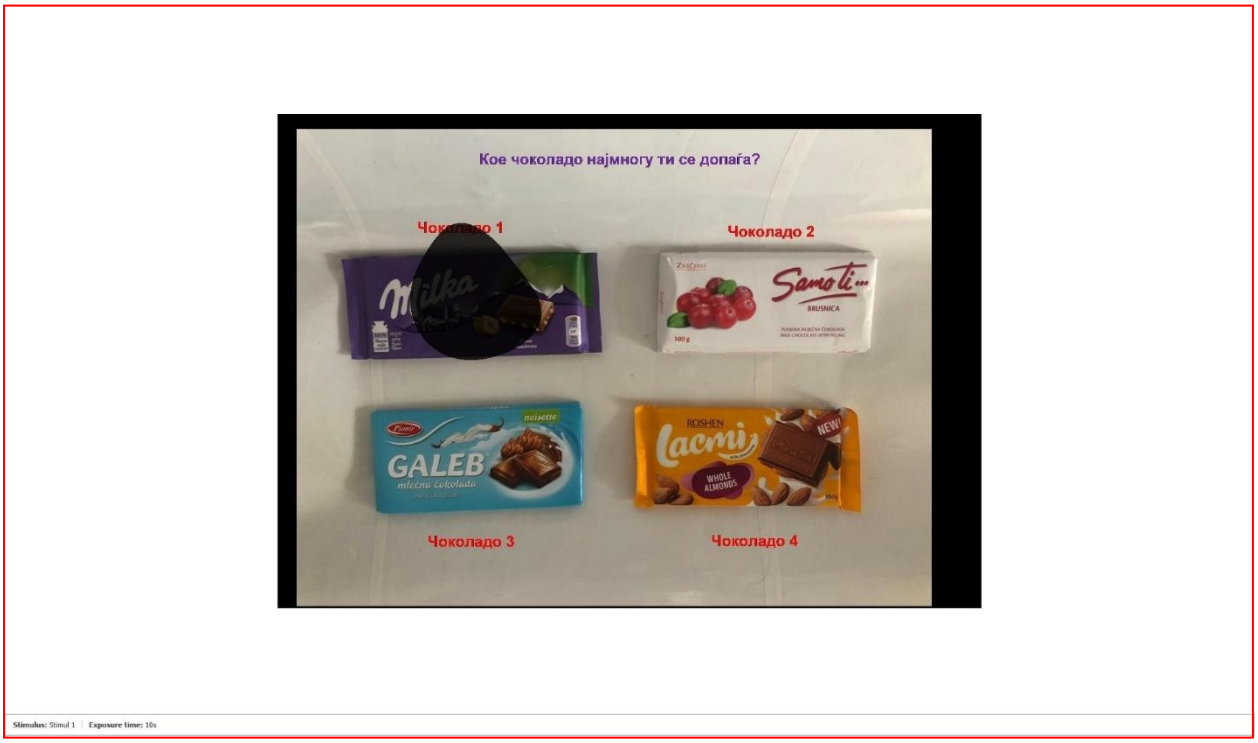


Слика 9. Топлинска мапа, стимул 1



Слика10 . Топлинска мапа, стимул 1

На слика 11 е прикажана збирна топлинска мапа за стимул 1 каде површината на сенката е пропорционална на нивото на интерес на сите испитаници заедно. Се забележува дека површината на сенката е најголема за чоколадо 1, додека за другите чоколади скоро и да не постои интерес.



Слика 11. Топлинска мапа, стимул 1

На слика 12 е прикажана збирна топлинска мапа за стимул 2 (од сите испитаници заедно). Кај овој стимул се добиени следните метрики со двата метода на истражување (Табели од споредување на резултати, на крајот на магистерската работа):

Анкетен прашалник:

- **Chipsy 1**: Два од 11 испитаници го избрале овој производ – 18.18%.
- **Chipsy 2**: Пет од 11 испитаници го избрале овој производ – **45.45%.**
- **Chipsy 3**: Четири од 11 испитаници го избрале овој производ – 36.36%.

Биосензорско мерење:

- **Chipsy 1**: 2.5 од 11 испитаници го избрале овој производ – 22.73%.

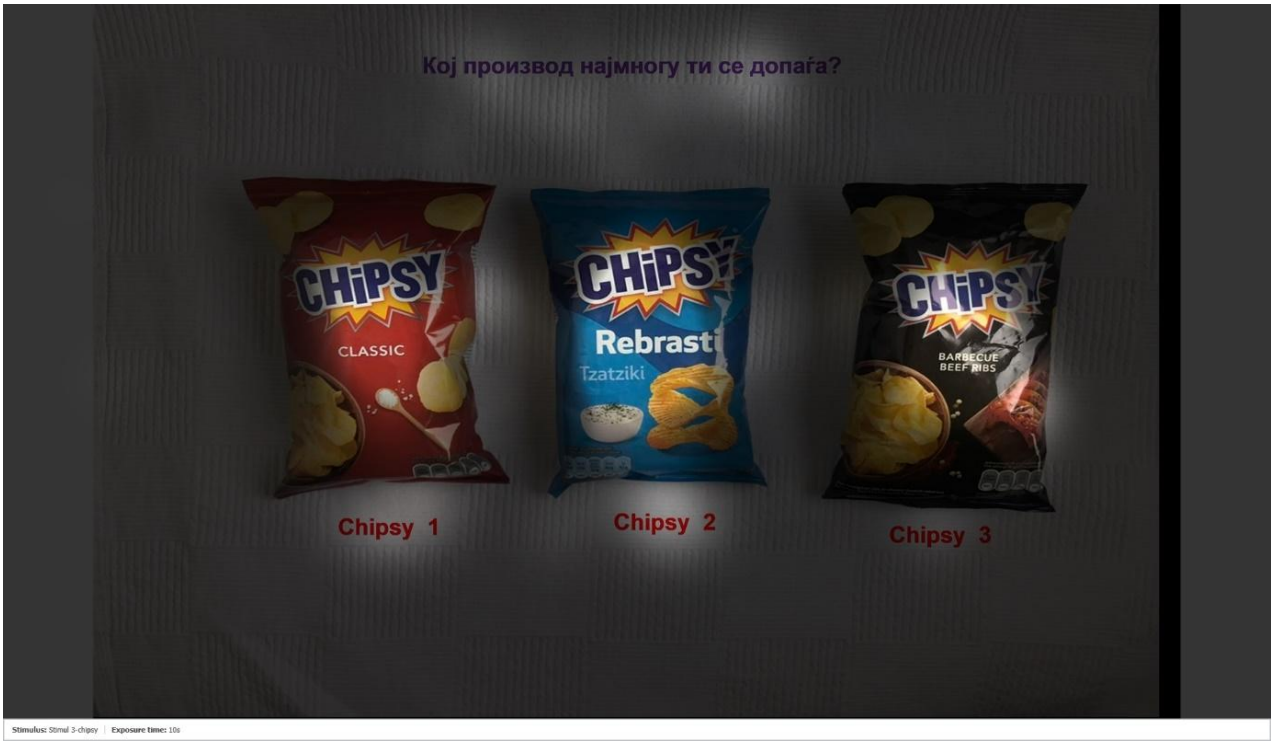
- **Chipsy 2**: 4.5 од 11 испитаници го избрале овој производ – **40.91%.**
- **Chipsy 3**: Четири од 11 испитаници го избрале овој производ – 36.36%.

Може да се забележи дека и во двете истражувања кај испитаниците, најголем интерес има за **Chipsy 2**. Но, исто така, нешто помал интерес има за **Chipsy 3**.



Слика 12. Топлинска мапа, стимул 2

На слика 13 и слика 14 се прикажани збирни топлински мапи за стимул 2 во два различни стилови на прикажување, каде како индикатор за фокусирано внимание се ниво на осветленост на одредени производи, сегменти од производи и површина на сенка генерирана на одреден производ. Овие индикатори даваат визуелен приказ за нивото на интерес за секој производ од сите испитаници заедно. Се забележува дека **Chipsy 2** има највисоко ниво на осветленост и најголема површина на сенка. Овој производ има најголем процент од сите испитаници (го избрале како производ кој најмногу им се допаѓа).



Слика 13. Топлинска мапа, стимул 2



Слика 14. Топлинска мапа, стимул 2

б)Низи на фиксации **(Fixation sequences)**

Низа на фиксации се процесира врз основа на:

- **Позиција на одредена фиксација** (**КАДЕ е насочен погледот**);
- **Информација за времето кога е реализиран одреден поглед** (**КОГА**).

Зависи од:

- **Каде** гледал секој испитаник во одреден момент на следење на стимулот?
- **Колку** време потрошил во гледање на одреден детал од стимулот?
- Дава увид во **редоследот на гледање**: Покажува каде секој испитаник погледнал прв, втор, трет пат, итн.

Оваа метрика најчесто се користи во истражување со биосензорски систем за следење на очи.

Низа од фиксации се реализира **за секој испитаник посебно** и **за секој стимул посебно**.



Слика 15: Низа на фиксации на испитаник **4** на стимул **1**

На слика 15 е прикажана низа од фиксации на испитаник 4, за стимул 1. Визуелниот приказ покажува дека овој испитаник својот поглед (и време поминато во гледање) најдолго го фокусирал на чоколадо 1 (тоа е негов избор и со двата методи на истражување (анкетен прашалник и биосензорско мерење). Исто така е прикажана дилатација (проширување на зеница) на двете очи на испитаникот. Проширувањето на зеницата е пропорционално на нивото на возбуда на испитаникот. Се забележува дека овој испитаник има зголемено ниво на возбуда кога го гледа својот омилен производ (од понудените четири). Исто така, прикажана е и фреквенција на трепкање при опсервација на стимулот (има две трепкања во период од 10 секунди на следење на екрански стимул на компјутер). Зголемена фреквенција на трепкање е пропорционална на намалено ниво на концентрација и зголемена поспаност.

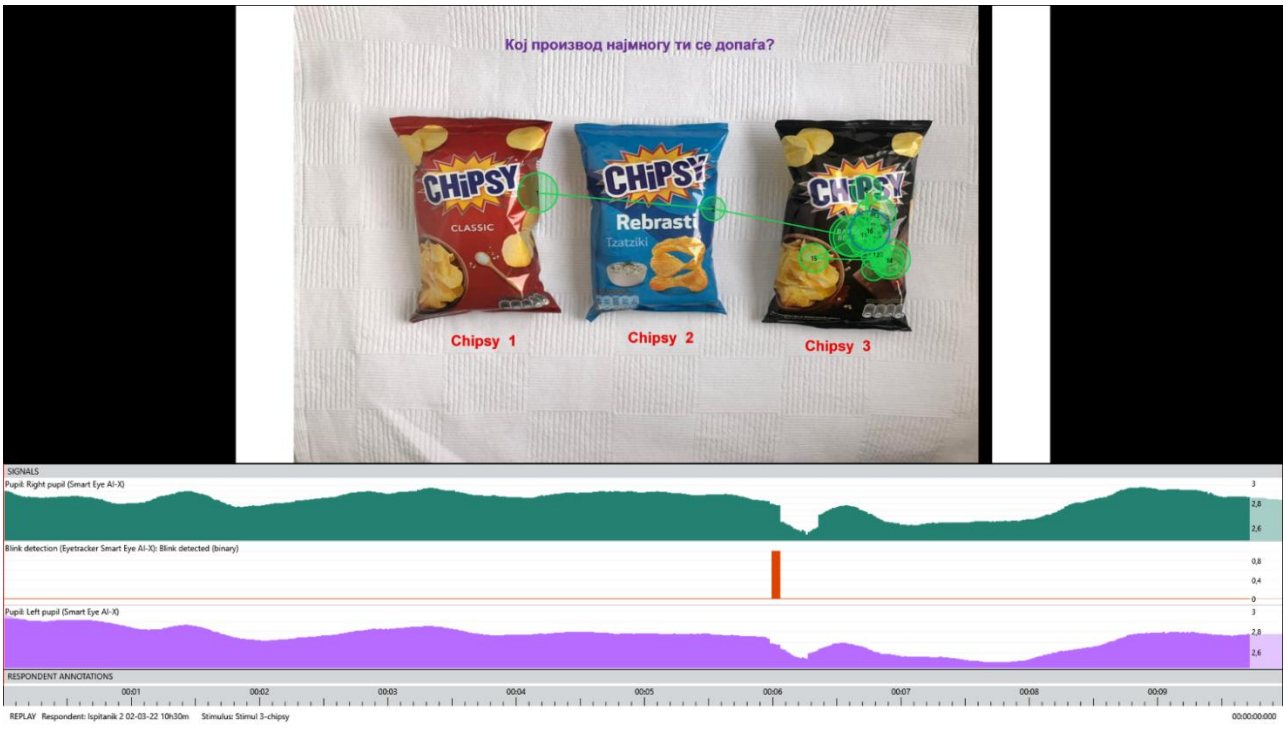


Слика 16: Низа на фиксации на испитаник **1** на стимул **1**

На слика 16 е прикажана низа од фиксации на испитаник 1 при следење на стимул 1. Се забележува дека овој испитаник покажал заинтересираност за повеќе производи, но најмногу за чоколадо 1 и чоколадо 4. Преку следење на дилата-

ција на зениците на двете очи се забележува дека овој испитаник зголемено ниво на возбуда покажал кога го следел чоколадо 1 (негов избор и со двата методи на истражување). Овој испитаник не направил ниту едно трепкање за период додека го следел стимулот на компјутерскиот екран, што покажува дека имал високо ниво на концентрација и фокусираност додека правел избор на производ кој најмногу му се допаѓа.

Слика 17 дава визуелен приказ на фиксации за испитаник 2 кога набљудува екрански стимул 2. Се забележува дека овој испитаник најголем интерес покажал за третиот производ (**Chipsy 3**), каде најголем дел од своето време на опсервација посветил на овој производ. За другите два производи не покажал никаков интерес. Прикажана е и дилатација на лева и десна зеница кои покажуваат кога овој испитаник имал зголемено ниво на возбуда. Во текот на следењето на екранскиот стимул, овој испитаник имал само едно трепнување, што покажува дека неговата концентрација и фокус биле на исклучително високо ниво.



Слика 17: Низа на фиксации на испитаник **2** на стимул **2**

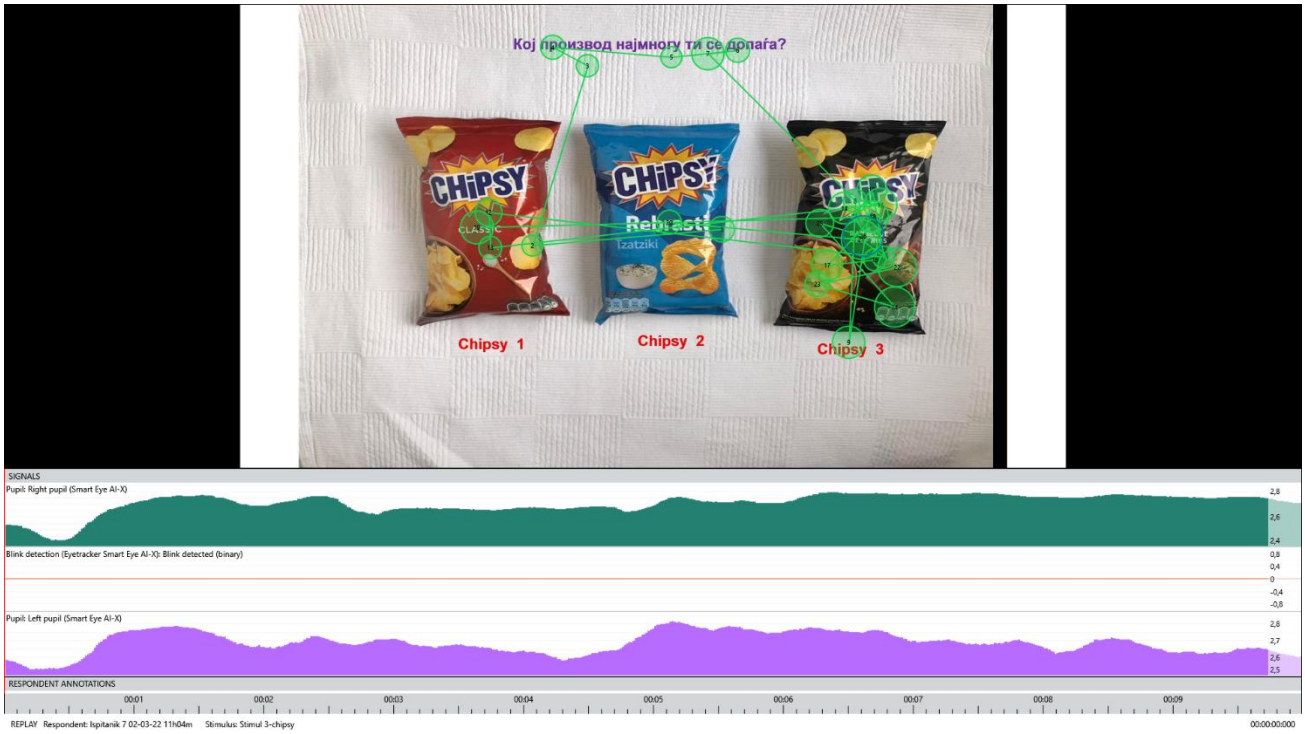
На слика 18 е даден визуелна слика на низа на фиксации на испитаник 4 при биосензорско следење на стимул 2. Се забележува дека најголем дел на фиксации се лоцирани на производот **Chipsy 2**, што покажува дека овој испитаник најголем интерес покажал за овој производ. Дилатацијата на зениците покажува одредено константно ниво на возбуда. Фреквенцијата на трепкање е висока (четири трепкања за период од 10 секунди следење на стимулот) што покажува намалено ниво на концентрација. На анкетниот прашалник овој испитаник го заокружил (избрал) **Chipsy 2** што покажува дека нема дискрепанца (отстапување) помеѓу двата методи на истражување.



Слика 18: Низа на фиксации на испитаник **4** на стимул **2**

На слика 19 е даден визуелна слика на низа на фиксации на испитаник 7 при биосензорско следење на стимул 2. Се забележува дека најголем дел на фиксации се лоцирани на производот **Chipsy 3**, што покажува дека овој испитаник најголем интерес покажал за овој производ. Фреквенцијата на трепкање е ниска (нема трепкања за период од 10 секунди следење на стимулот) што покажува зголемено ниво на концентрација. На анкетниот прашалник овој испитаник го

заокружил (избрал) **Chipsy 3** што покажува дека нема дискрепанца (отстапување) помеѓу двата методи на истражување.



Слика 19: Низа на фиксации на испитаник **7** на стимул **2**

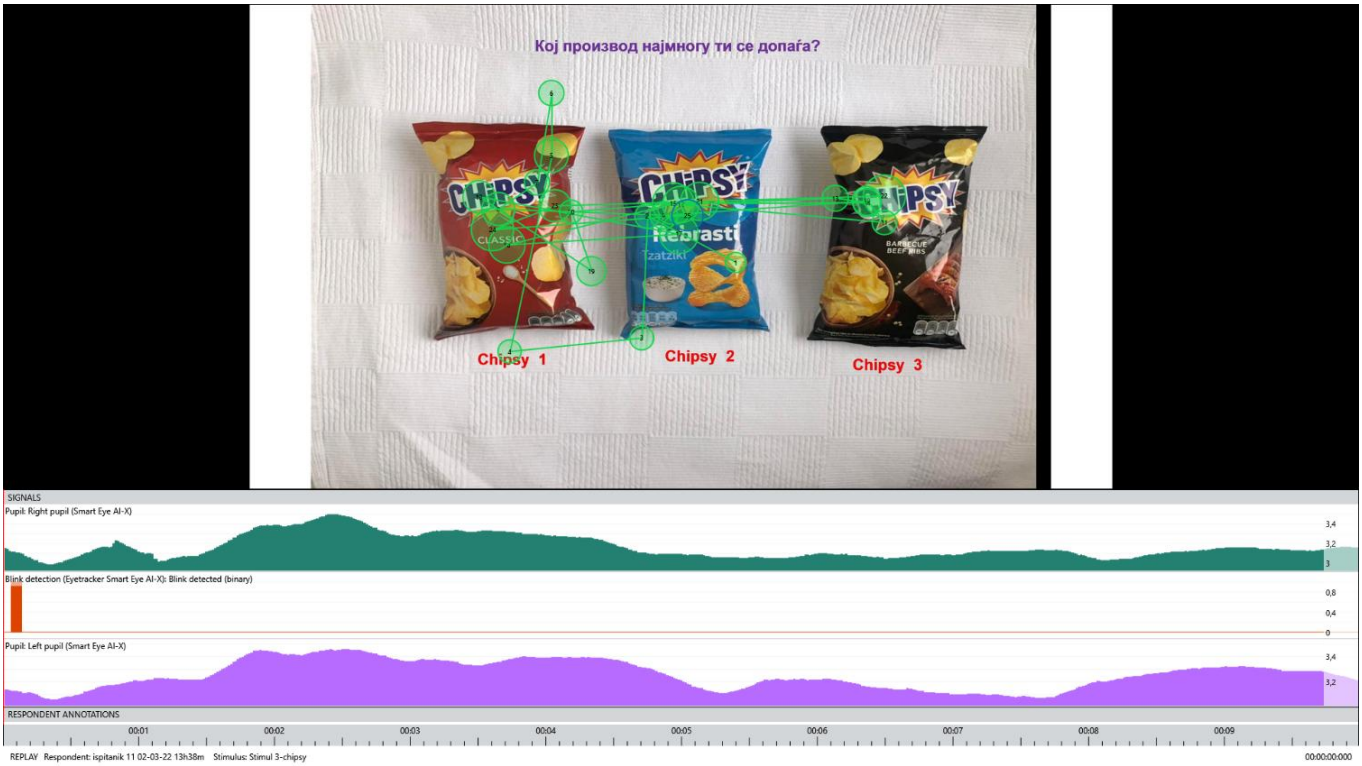
На слика 20 е даден визуелен приказ на низа на фиксации на испитаник 8 при биосензорско следење на стимул 2. Се гледа дека фиксациите се распределени на трите производи (испитаникот се двоуми кој производ да го избере – нема омилен). Со детална анализа на резултати од биосензорско следење на очите се добиени следните резултати за процент на фокусирано внимание:

- За **Chipsy 2** изнесува **37%**;
- За **Chipsy 3** изнесува **24.9%.** Изборот е помеѓу овие два производи.

На анкетниот прашалник испитаникот го заокружил **Chipsy 2.** Од анкетниот прашалник не може да се открие дилемата што ја има потсвесното размислување – избор помеѓу **Chipsy 2** и **Chipsy 3**. Ова покажува дека при потсвесно одлучување (брзо одлучување) испитаникот ќе го избере еден од овие два производи. Но

при рационално (бавно) размислување, испитаникот недвосмислено ќе се одлучи за **Chipsy 2**.

Исто така, следењето на дилатацијата на зениците покажува големи скокови во нивото на возбуда. При детална анализа се открива дека вој испитаник зголемено ниво на возбуда покажува при следење на **Chipsy 2.**



Слика 20: Низа на фиксации на испитаник **8** на стимул **2**

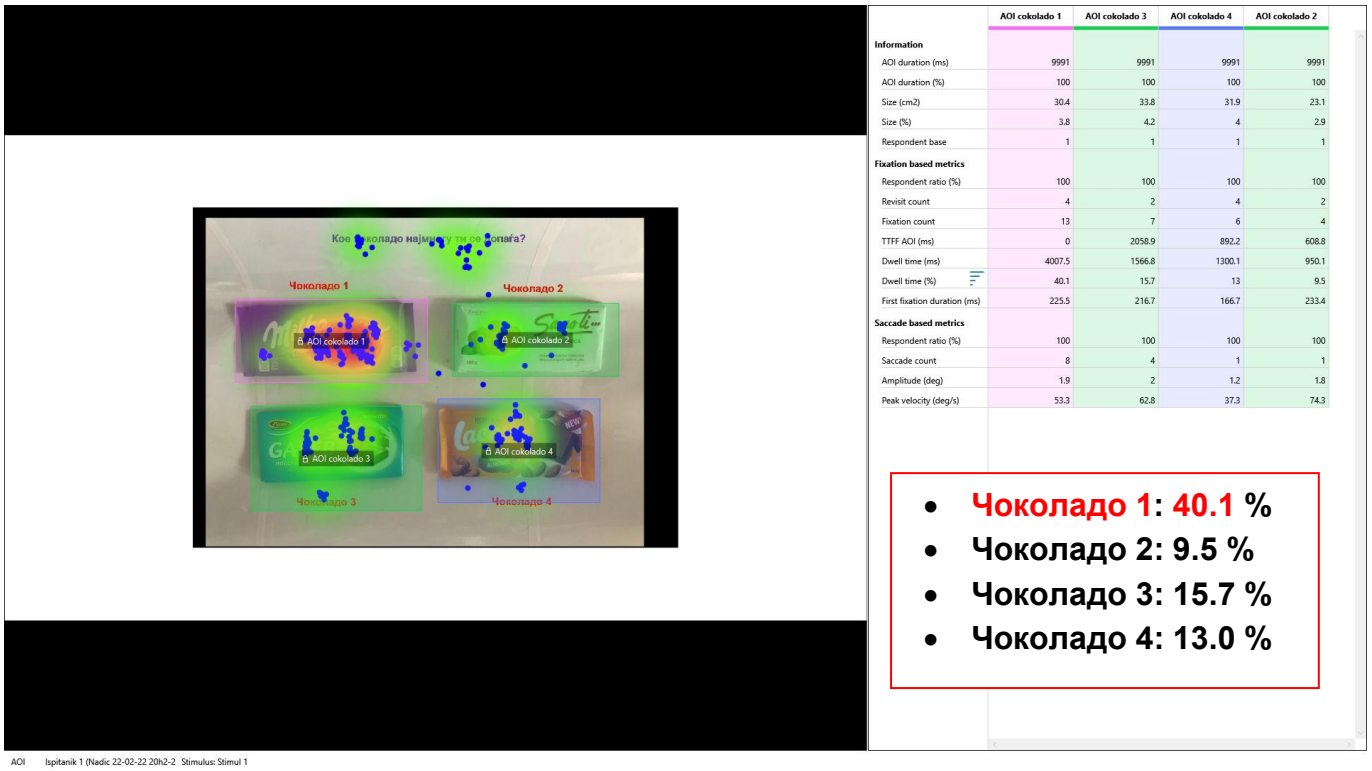
Низа на фиксации се прикажува:

- **Динамички** (движење на фиксации како што е креирана – реконструкција како е реализирано следењето на стимулот од испитаникот);
- **Статички** (се покажуваат сите фиксации) ;
- **Рачно** (детално се гледа секоја фиксација кога се појавила, колку таа траела и каде се создала).

б)Области на интерес (Areas of interest): Поединечни анализи

Области на интерес претставуваат кориснички дефинирани **бафер- зони** на прикажан стимул. Тие прикажуваат метрики за одделни **AOI** со што се овозможува да се реализира **определување на перформанси** на две или повеќе специфични области на даден стимул. Ова се користи за да се реализираат **СПОРЕДУВАЊА** и други **статистички** анализи на иста сцена.

Со креирање на одредена област од интерес, се покажуваат поголем број на пресметувања кои даваат карактеристики за визуелното внимание за одреден стимул.



	AOI cokolado 1	AOI cokolado 3	AOI cokolado 4	AOI cokolado 2
Information				
AOI duration (ms)	9991	9991	9991	9991
AOI duration (%)	100	100	100	100
Size (cm2)	30.4	33.8	31.9	23.1
Size (%)	3.8	4.2	4	2.9
Respondent base	1	1	1	1
Fixation based metrics				
Respondent ratio (%)	100	100	100	100
Revisit count	4	2	4	2
Fixation count	13	7	6	4
TTFF AOI (ms)	0	2058.9	892.2	608.8
Dwell time (ms)	4007.5	1566.8	1300.1	950.1
Dwell time (%)	40.1	15.7	13	9.5
First fixation duration (ms)	225.5	216.7	166.7	233.4
Saccade based metrics				
Respondent ratio (%)	100	100	100	100
Saccade count	8	4	1	1
Amplitude (deg)	1.9	2	1.2	1.8
Peak velocity (deg/s)	53.3	62.8	37.3	74.3

- **Чоколадо 1: 40.1 %**
- **Чоколадо 2: 9.5 %**
- **Чоколадо 3: 15.7 %**
- **Чоколадо 4: 13.0 %**

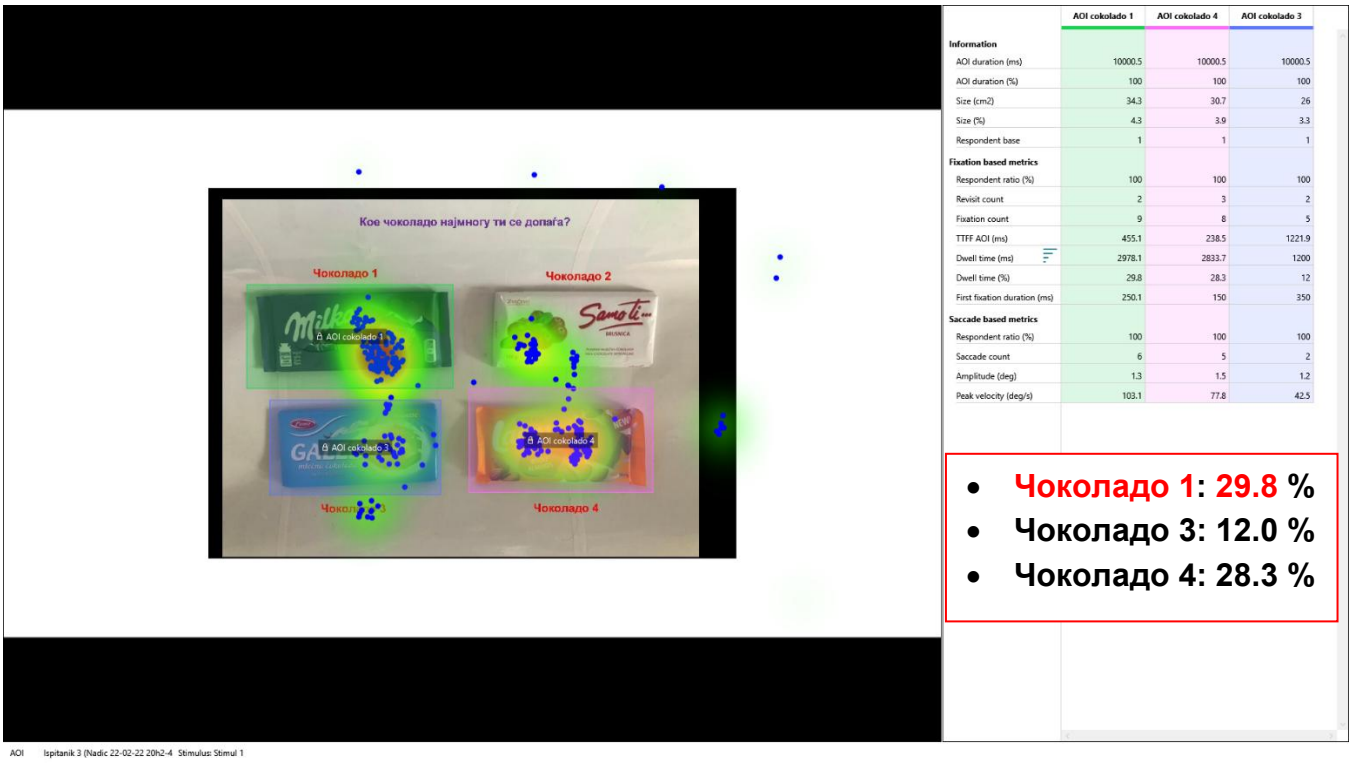
Слика 21. Област на интерес, стимул 1, испитаник 1

На слика 21 е прикажана област на интерес на стимул 1, за испитаник 1. Генерирани се четири (4) бафер- зони посебно околу секој производ. Прикажана е топлинска мапа и поединечни фиксации (грозд) кои се прикажани со сини точки. Се забележува дека најголем интерес овој испитаник покажал за чоколадо 1 каде

најдолго време го задржал неговиот поглед. Со метриката доживеано време на следење (Dwell time %) се покажува нивото на интерес за секој производ:

- Чоколадо 1: 40.1%;
- Чоколадо 3: 15.7%;
- Чоколадо 4: 13.0%;
- Чоколадо 2: 9.5%.

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Чоколадо 1 (40.1%), а значително помал интерес за другите производи.

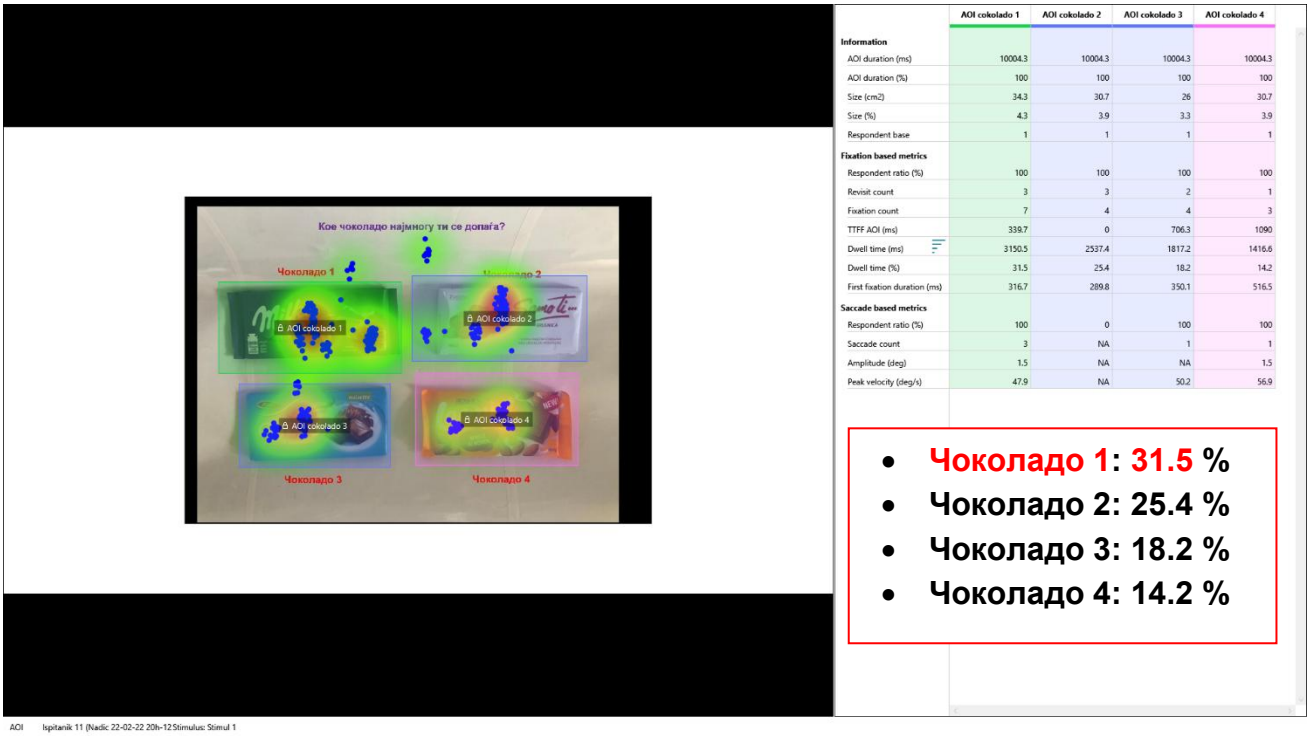


Слика 22. Област на интерес, стимул 1, испитаник 3

На слика 22 е прикажана област на интерес на стимул 1, за испитаник 3. Генерирани се три (3) бафер- зони посебно околу три производа кои биле од најголем интерес за испитаникот. Прикажана е топлинска мапа и поединечни фиксации (грозд) кои се прикажани со сини точки. Се забележува дека најголем интерес овој испитаник покажал за чоколадо 1 (29.8%), но незначително помал интерес за чоколадо 4 (28.3%). Со метриката доживеано време на следење (Dwell time %) се покажува нивото на интерес за секој производ:

- Чоколадо 1: 29.8%;
- Чоколадо 4: 28.3%;
- Чоколадо 3: 12.0%;

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Чоколадо 1 и за чоколадо 4, а значително помал интерес за чоколадо 3. Ова покажува дека овој испитаник не може да се одлучи помеѓу овие два производи. На анкетен прашалник овој испитаник го заокружил чоколадо 1, додека биосензорското мерење на потсвесно размислување покажува неодлучност помеѓу чоколадо 1 и чоколадо 4. Ова во практика значи дека, ако од испитаникот се бара брзо да се одлучи (потсвесно размислување) за одреден производ од стимул 1, тој ќе се двоуми помеѓу чоколадо 1 и чоколадо 4 (може да избере било кој од овие производи). Но кога би имал подолго време за размислување (рационално размислување) тој одлучно би избрал чоколадо 1 (Табела 1, за стимул 1).



	AOI cokolado 1	AOI cokolado 2	AOI cokolado 3	AOI cokolado 4
Information				
AOI duration (ms)	10004.3	10004.3	10004.3	10004.3
AOI duration (%)	100	100	100	100
Size (cm2)	34.3	30.7	26	30.7
Size (%)	4.3	3.9	3.3	3.9
Respondent base	1	1	1	1
Fixation based metrics				
Respondent ratio (%)	100	100	100	100
Revisit count	3	3	2	1
Fixation count	7	4	4	3
TTFF AOI (ms)	339.7	0	706.3	1090
Dwell time (ms)	3150.5	2537.4	1817.2	1416.6
Dwell time (%)	31.5	25.4	18.2	14.2
First fixation duration (ms)	316.7	289.8	350.1	516.5
Saccade based metrics				
Respondent ratio (%)	100	0	100	100
Saccade count	3	NA	1	1
Amplitude (deg)	1.5	NA	NA	1.5
Peak velocity (deg/s)	47.9	NA	50.2	56.9

Слика 23. Област на интерес, стимул 1, испитаник 11

На слика 23 е прикажана област на интерес на стимул 1, за испитаник 11. Генерирани се четири (4) бафер зони посебно околу секој производ. Прикажана е топлинска мапа и поединечни фиксации (грозд) кои се прикажани со сини точки. Се забележува дека најголем интерес овој испитаник покажал за чоколадо 1 (31.5%), но незначително помал интерес за чоколадо 2 (25.4%). Со метриката доживеано време на следење (Dwell time %) се покажува нивото на интерес за секој производ:

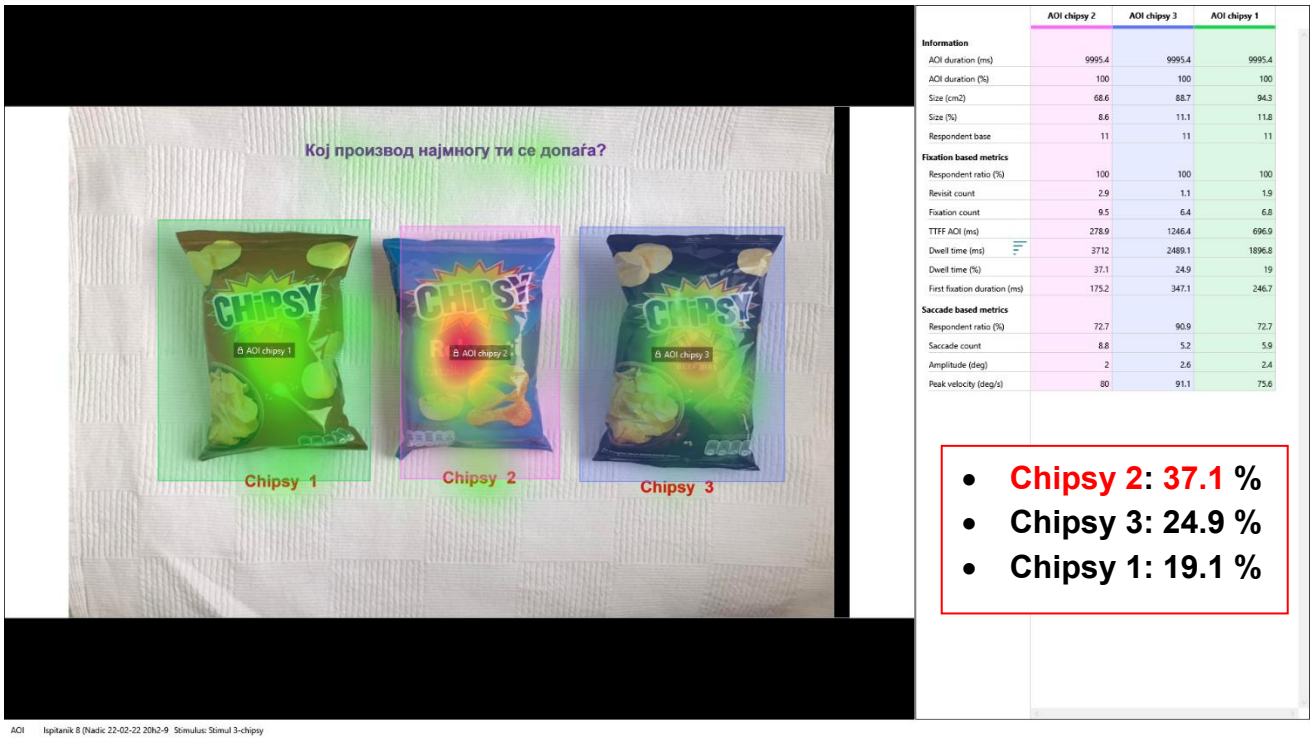
- **Чоколадо 1**: 31.5%;
- Чоколадо 2: 25.4%;
- Чоколадо 3: 18.2%;
- Чоколадо 4: 14.2%.

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Чоколадо 1 и за чоколадо 2, а значително помал интерес за чоколадо 3 и 4. Ова покажува дека овој испитаник не може да се одлучи помеѓу овие два производи (1 и 2). На анкетен прашалник овој испитаник го заокружил чоколадо 1, додека био-сензорското мерење на потсвесно размислување покажува неодлучност помеѓу чоколадо 1 и чоколадо 2. Ова во практика значи дека, ако од испитаникот се бара брзо да се одлучи (потсвесно размислување) за одреден производ од стимул 1, тој ќе се двоуми помеѓу чоколадо 1 и чоколадо 2 (може да избере било кој од овие производи, но со поголема веројатност за чоколадо 1). Но кога би имал подолго време за размислување (рационално размислување) тој одлучно би избрал чоколадо 1 (Табела 1, за стимул 1).

На слика 24 е прикажана област на интерес на стимул 2, за испитаник 8. Генерирани се три (3) бафер- зони оделно околу секој производ. Прикажана е топлинска мапа. Се забележува дека најголем интерес овој испитаник покажал за Chipsy 2 каде најдолго време го задржал неговиот поглед. Со метрика доживеано време на следење на поглед на испитаникот (Dwell time %) се покажува нивото на интерес за секој производ:

- **Chipsy** 2: 37.1%
- Chipsy 3: 24.9%
- Chipsy 1: 19.0%

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Chipsy 2 (37.1%), а значително помал интерес за другите производи.



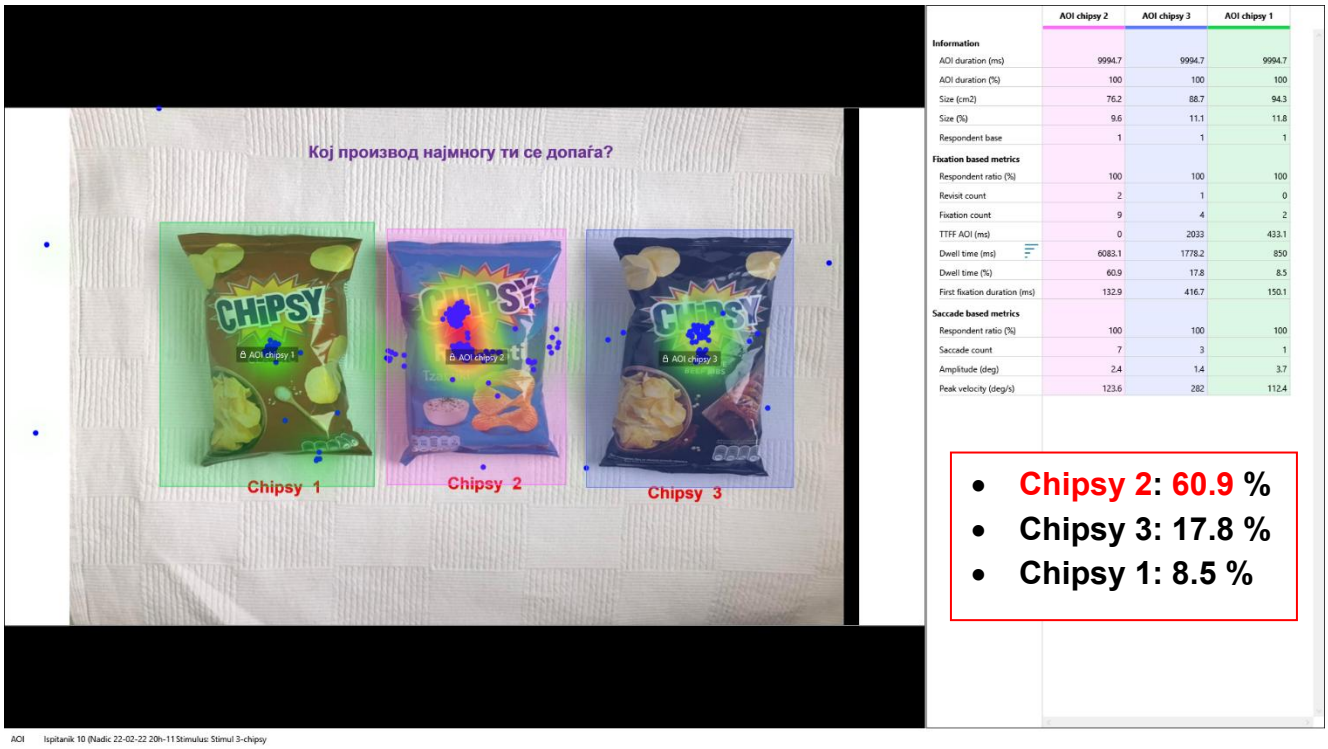
Слика 24. Област на интерес, стимул 2, испитаник 8

На слика 25 е прикажана област на интерес на стимул 2, за испитаник 10. Генерирани се три (3) бафер-зони одделно околу секој производ. Прикажана е топлинска мапа и поединечни фиксации (грозд) кои се прикажани со сини точки. Се забележува дека најголем интерес овој испитаник покажал за Chipsy 2 каде најдолго време го задржал неговиот поглед (60.9%). Со метрика доживеано време на следење на поглед на испитаникот (Dwell time %) се покажува нивото на интерес за секој производ:

- **Chipsy** 2: 60.9%;
- Chipsy 3: 17.8%;
- Chipsy 1: 8.5%.

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Chipsy 2 (60.9%), а значително помал интерес за другите производи. На анкетниот

прашалник овој испитаник го заокружил Chipsy 2 што покажува дека не постои дискрепанца (отстапување) помеѓу двата методи на истражување: биосензорско следење на очи (потсвесно размислување) и анкетно истражување (рационално размислување).

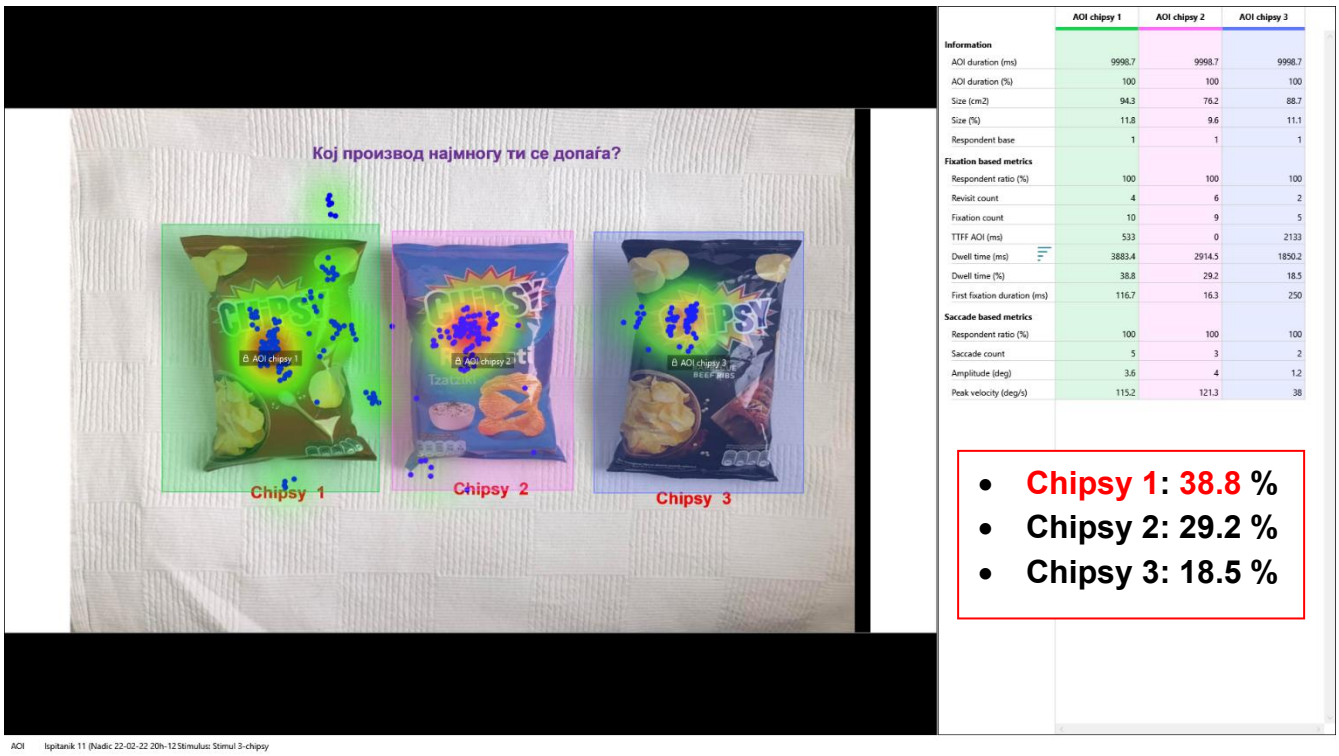


Слика 25. Област на интерес, стимул 2, испитаник 10

На слика 26 е прикажана област на интерес на стимул 2, за испитаник 11. Генерирани се три (3) бафер- зони оделно околу три производ кои биле од најголем интерес за испитаникот. Прикажана е топлинска мапа и поединечни фиксации (грозд) кои се прикажани со сини точки. Се забележува дека најголем интерес овој испитаник покажал за Chipsy 1 (38.8%), но незначително помал интерес за Chipsy 2 (29.2%). Со метриката доживеано време на следење (Dwell time %) се покажува нивото на интерес за секој производ:

- **Chipsy 1**: **38.8%**
- Chipsy 2: 29.2%
- Chipsy 3: 18.5%

Оваа метрика покажува дека овој испитаник најголем интерес покажал за Chipsy 1 и за Chipsy 2, а значително помал интерес за Chipsy 3. Ова покажува дека во прашалнико овој испитаник го заокружил Chipsy 2, додека биосензорското мерење кај овој испитаник не може да се одлучи помеѓу овие два производи. На анкетен прашалник потсвесно размислување покажува неодлучност помеѓу Chipsy 1 и Chipsy 2. Ова во практика значи дека, ако од испитаникот се бара брзо да се одлучи (пот-свесно размислување) за одреден производ од стимул 2, тој ќе се двоуми помеѓу Chipsy 1 и Chipsy 2 (може да избере било кој од овие производи, со поголема ве-ројатност да биде избран Chipsy 1). Но, кога би имал подолго време за размислување (рационално размислување) тој одлучно би избрал Chipsy 2 (Табела 2, за стимул 2). Ова покажува дека постои дискрепанца (отстапување) помеѓу двата методи на истражување. При биосензорско (брзо размислување) мерење испитаникот би го избрал **Chipsy 1**, а при анкетно истражување (бавно, рационално размислување) испитаникот би го одбрал **Chipsy 2**.



	AOI chipsy 1	AOI chipsy 2	AOI chipsy 3
Information			
AOI duration (ms)	9998.7	9998.7	9998.7
AOI duration (%)	100	100	100
Size (cm2)	94.3	76.2	88.7
Size (%)	11.8	9.6	11.1
Respondent base	1	1	1
Fixation based metrics			
Respondent ratio (%)	100	100	100
Revisit count	4	6	2
Fixation count	10	9	5
TTFF AOI (ms)	533	0	2133
Dwell time (ms)	3883.4	2914.5	1850.2
Dwell time (%)	38.8	29.2	18.5
First fixation duration (ms)	116.7	16.3	250
Saccade based metrics			
Respondent ratio (%)	100	100	100
Saccade count	5	3	2
Amplitude (deg)	3.6	4	1.2
Peak velocity (deg/s)	115.2	121.3	38

AOI Ispitanik 11 (Nadic 22-02-22 20h-12 Stimulus: Stimul 3-chipsy

Слика 26. Област на интерес, стимул 2, испитаник 11

е) Области на интерес (Areas of interest): **збирни** анализи

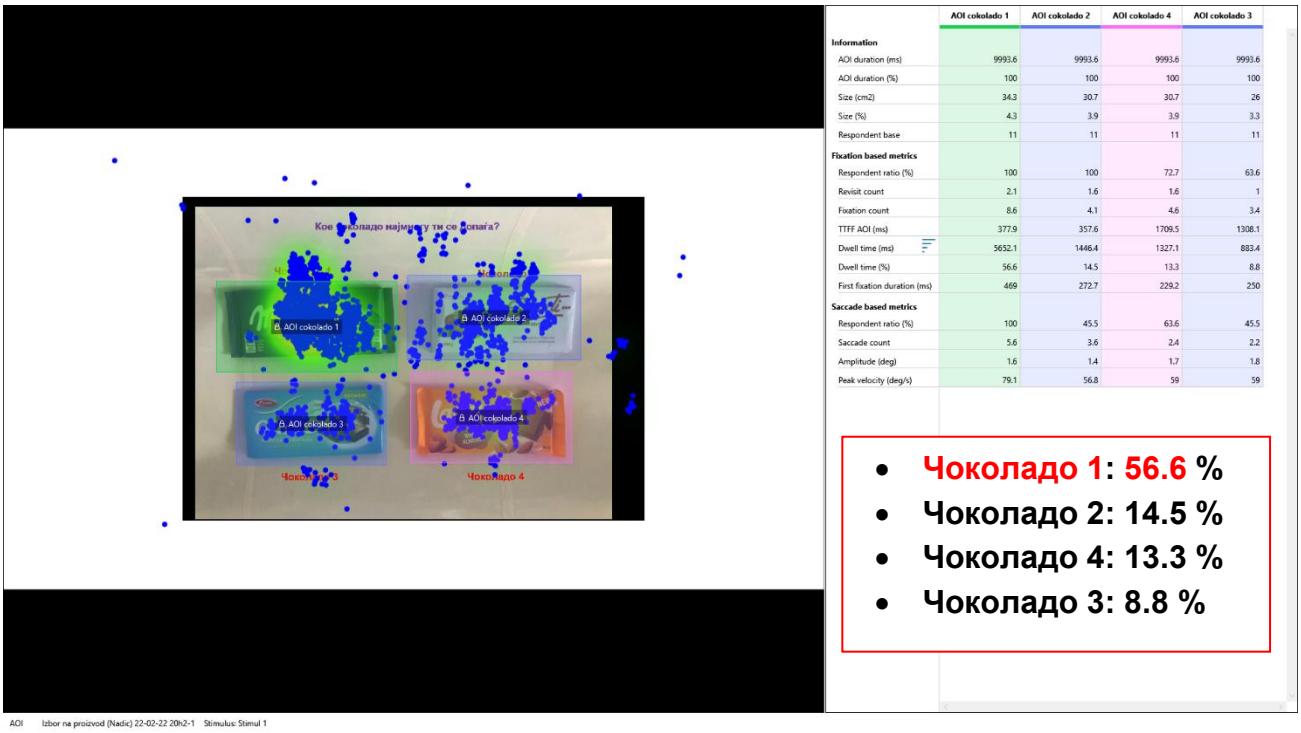
Збирна анализа прикажува збирни резултати (мерења) од сите испитаници заедно и најмногу се користи во реални истражувања бидејќи покажува тренд на одредена целна група, а на корисниците (маркетинг- менаџери, психолози, педагози и многу други) им овозможува да видат како одредена целна група одговара на одредено прашање.

На слика 27 е прикажан збирен графички приказ за стимул 1. Бафер- зоните (генерирани околу секој производ) покажуваат дека на испитаниците најмногу им се допаднал првиот производ (**56.6%**) (чоколадо 1), додека значајно помал интерес има кон другите производи. На Анкетниот прашалник, **90.9%** од испитаниците го заокружиле чоколадо 1 (Табела 1).



Слика 27. Област на интерес, стимул 1

На слика 28 е прикажана иста збирна топлинска мапа како и слика 27, но со прикажани поединечни фиксации (грозд). Се забележува дека најмногу фиксации има на првиот производ (чоколадо 1) – испитаниците најдолго својот поглед го задржале на овој сегмент од стимулот.



Слика 28. Област на интерес, стимул 1

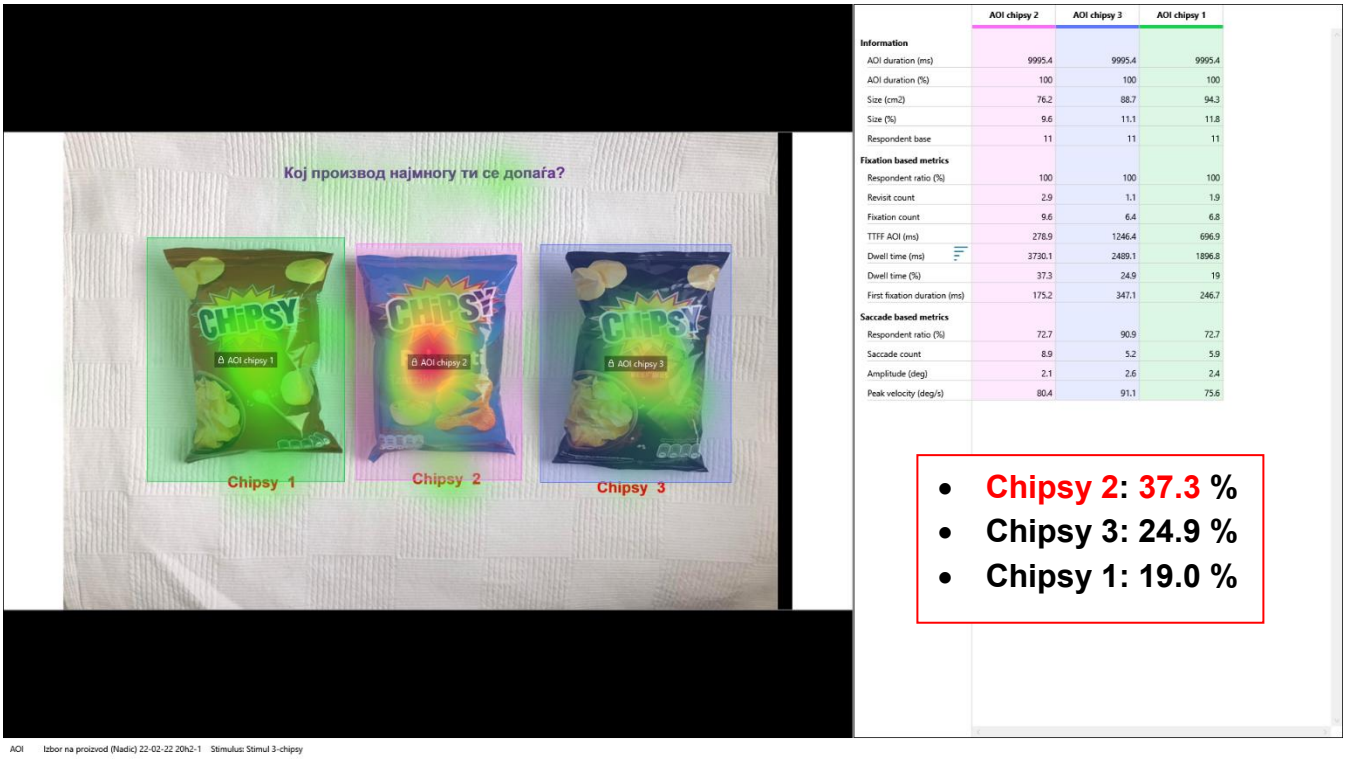
На слика 29 е прикажан збирен графички приказ за стимул 2. Бафер- зоните (генерирани околу секој производ) покажуваат дека на испитаниците најмногу им се допаднал вториот производ (**37.3%**) (**Chipsy 2**), додека нешто помал интерес има кон другите производи. На Анкетниот прашалник,

40.91% од испитаниците го заокружиле **Chipsy 2** (Табела 1);

За **Chipsy 3** се определиле **36.36%** од испитаниците;

За **Chipsy 1** се определиле **22.73%** од испитаниците.

Споредувајќи ги резултатите од двата методи на истражување, може да се заклучи дека крајните резултати се во корелација, даваат приближно слични одговори. Ова покажува дека биосензорската технологија (биосензор за следење на очи) е сигурен метод за добивање релевантни резултати. И уште повеќе, биосензорското истражување е ослободено од пристрасност на испитаниците, за разлика од анкетно истражување каде може да се појави пристрасност кај одредени испитаници од различни причини.



Слика 29. Област на интерес, стимул 2

1. **Споредување на поединечни вредности (посебно за секој стимул) за сите испитаници – Табела 1**

Стимул 1			
Испитаник	**Анкетен прашалник**	**Биометриско мерење со Eye Tracker**	**Дискрепанца (2-3)**
Испитаник **1**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **2**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **3**	**Чоколадо 1**	**Чоколадо 1 (4)**	**има: 1 и 4**
Испитаник **4**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **5**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **6**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **7**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **8**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **9**	**Чоколадо 2**	**Чоколадо 2**	**нема**
Испитаник **10**	**Чоколадо 1**	**Чоколадо 1**	**нема**
Испитаник **11**	**Чоколадо 1**	**Чоколадо 1 (2)** **31.5% (25.4%)**	**има: 1 и 2**
Збирни резултати	**Чоколадо 1: 10/11 (90.9%) Чоколадо 2: 1/11 (9.1%)**	**Чоколадо 1: 9/11 (81.8%) Чоколадо 2: 1.5/11 (13.63%) Чоколадо 4: 0.5/11 (4.54%)**	**Чоколадо 1: (90.9:81.8)% Чоколадо 2: (9.1:13.63)% Чоколадо 4: (0:4.54)%**

Стимул 2 (Табела 2)			
Испитаник	**Анкетен прашалник**	**Биометриско мерење со Eye Tracker**	**Дискрепанца (2-3)**
Испитаник **1**	**Chipsy 3**	**Chipsy 3**	**нема**
Испитаник **2**	**Chipsy 3**	**Chipsy 3**	**нема**
Испитаник **3**	**Chipsy 1**	**Chipsy 1**	**нема**
Испитаник **4**	**Chipsy 2**	**Chipsy 2**	**нема**
Испитаник **5**	**Chipsy 2**	**Chipsy 2**	**нема**
Испитаник **6**	**Chipsy 1**	**Chipsy 1**	**нема**
Испитаник **7**	**Chipsy 3**	**Chipsy 3**	**нема**
Испитаник **8**	**Chipsy 2**	**Chipsy 2 (3)** **37.0% (24.9%)**	**има: 2 и 3**
Испитаник **9**	**Chipsy 2**	**Chipsy 2**	**нема**
Испитаник **10**	**Chipsy 3**	**Chipsy 2 (3)** **60.9% (17.8%)**	**има: 2 и 3**
Испитаник **11**	**Chipsy 2**	**Chipsy 1 (2)** **38.8% (29.2%)**	**има: 1 и 2**
Збирни резултати	**Chipsy 3: 4/11 (36.36%)** **Chipsy 2: 5/11 (45.45%)** **Chipsy 1: 2/11 (18.18%)**	**Chipsy 3: 4/11 (36.36%))** **Chipsy 2: 4.5/11 (40.91%)** **Chipsy 1: 2.5/11 (22.73%)**	**Chipsy 3: (36.36: 36.36)%** **Chipsy 2: (45.45:40.91)%** **Chipsy 4: (18.18:22.73)%**



Заклучни согледувања

Гласовна анализа се користи во многу апликации, но нејзината вредност е особено добро докажана во контекст на клинички истражувања. На пример, гласовната анализа може да идентификува биомаркери за Паркинсонова болест, депресија и малигни полипи на грлото. Гласовната анализа сé повеќе се користи во општествените и бихејвиоралните науки каде што може да открие промени во емоционални состојби. Анализата на гласот денес се движи во нови истражувачки полиња како што се автомобилската индустрија, човечките фактори и науката за перформансите.

Гласовна анализа е моќна алатка која обезбедува увид во човечкото однесување и психологија. Анализирањето на аспектите како тон и ритам помага да се разберат емоционалните состојби и благосостојбата. Тој е разновиден, неинвазивен и има ниска бариера за влез, што го прави привлечен за истражувачите.

Гласовна анализа (Voice analysis) се користи во многу различни видови истражувања и можеме да размислуваме за користење на гласовна анализа врз основа на видот на полето за истражување или типот на поставеното истражување. Во магистерскиот труд прво давам примери за тоа како гласовната анализа се применува во три различни истражувачки области, а потоа и како гласовната анализа се користи со различни истражувачки дизајни.

Клиничкото истражување со децении користеше гласовна анализа за да направи разлика помеѓу клиничката популација. На пример, студиите идентификуваа гласовни биомаркери за дизартричен говор, депресија, Алцхајмерова болест и Паркинсонова болест. Во клиничките истражувања, гласовната анализа се користи и за да се процени ефикасноста на третманот. На пример, со студии кои ја оценуваат ефикасноста на третманот за социјална анксиозност со помош на гласовна анализа (Laukka et al., 2008).

Во клиничкото истражување, гласовната анализа обично се објавува како дел од високо контролирани, лабораториски истражувачки студии каде што учесникот извршува говорна задача додека е соочен со компјутер или како дел од клиничко интервју.

Гласовната анализа се користи во различни истражувачки полиња, вклучувајќи клиничко истражување, социјални и бихејвиорални истражувања и маркетинг и деловни истражувања. Ваква анализа може да се примени на збирки податоци собрани со различни дизајни на студии, вклучувајќи тестови со гласно размислување, дијадични интеракции, презентации и интервјуа.

Откако гласовната анализа ќе ја заврши својата обработка, можеме да го отвориме прозорецот за повторување (the replay window) за еден од нашите испитаници и ќе можеме да ги визуелизираме гласовните сигнали со нивно избирање во левиот панел. Гласовните сигнали автоматски се синхронизираат и со аудио/видео снимањето и со сите други сигнали што можеби сме ги собрале за време на реализирање на снимањето.

Интеграцијата на вештачката интелигенција (AI - artificial intelligence) направи револуција во различни индустрии, а денес ја трансформира областа на истражувањето на човековото однесување. iMotions во 2023 година го интегрира API за софистицираното препознавање и анализа на говор на AssemblyAI во сопствената платформа iMotions Lab, за да може да им понуди анализа од говор во текст на своите корисници.

Оваа интеграција означува значајна пресвртница во научните и истражувачки напори за собирање податоци и анализа, овозможувајќи им на корисниците да отклучат подлабоки сознанија од говорниот јазик и да ги поттикнат истражувачите и аналитичарите со зголемени способности за разбирање и толкување на човечката комуникација.

Новата функција за анализа на говор во текст им овозможува на корисниците да импортираат видеа или аудио датотеки и преку API на AssemblyAI, аудиото автоматски да се транскрибира и анализира. Анализата на звукот вклучува откривање на број на говорници, дискурзивно откривање на валентност (употреба

на позитивни, негативни или неутрално оптоварени зборови) и сумирање на говорот.

Компонентите за анализа на говор во текст и говор се меѓу врвните карактеристики што голем број клиенти и корисници ги бараа низ годините. **Вредноста на анализирање интервјуа**, **фокус групи**, **говори** или **интеракции снимени на аудио или видео** не може да се потцени. Може да даде нова длабочина и да обезбеди анализа на речиси секоја снимена вербална секвенца.

Гласовна анализа (анализа на говор) е проучување на вокалните карактеристики и обрасци во говорната комуникација. Вклучува испитување на елементи како што се **тон**, **стапка на говор** и **емоционални знаци** за да се добие увид во аспекти како лингвистика, емоции и индивидуални особини. Оваа технологија има примена во области како што се **лингвистиката**, **психологија** и **здравство**.

Во практичното истражување на овој магистерски труд е користена биосензорска современа технологија за следење на очи на испитаници. Очите се прозорец на мозокот на испитаниците и покажуваат како мозокот размислува. Вакви истражувања денес користат многу компании во светот кои настојуваат да изградат современа маркетинг стратегија врз основа на размислувањата на потрошувачите.

Прилог: Анкетен прашалник.

Анкетен прашалник за избор на ХРАНА

АНКЕТЕН ПРАШАЛНИК (self-report)

Испитаник:

Во два дадени Тестови, заокружи го оној производ што најмногу ти се допаѓа:

ТЕСТ 1



ТЕСТ 2



Користена литература

1. Banai, I. P., Banai, B., & Bovan, K. (2017). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, 38(3), 309-314.
2. Bocklet, T., Steidl, S., Nöth, E., & Skodda, S. (2013, September). Automatic evaluation of parkinson's speech-acoustic, prosodic and voice related cues. In *Interspeech* (pp. 1149-1153).
3. Burkhardt, F., Wagner, J., Wierstorf, H., Eyben, F., & Schuller, B. (2023). Speech-based Age and Gender Prediction with Transformers. *arXiv preprint arXiv:2306.16962*.
4. Eyben, F. (2015). *Real-time speech and music classification by large audio feature space extraction*. Springer.
5. Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACMinternational conference on Multimedia* (pp. 1459-1462).
6. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong,
7. K. P. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE transactions on affective computing*, 7(2), 190-202.
8. Fan, M., Lin, J., Chung, C., & Truong, K. N. (2019). Concurrent think-aloud verbalizations and usability problems. *ACM Transactions on Computer-Human Interaction (TOCHI), 26(5),* 1-35.
9. Fan, M., Zhao, Q., & Tibdewal, V. (2021, May). Older adults' think-aloud verbalizations and speech features for identifying user experience problems. In

Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

10. Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of personality and social psychology*, 70(6), 1231.

11. Han, E. Y., Yun, J. Y., Chong, H. J., & Choi, K. G. (2018). Individual therapeutic singing program for vocal quality and depression in Parkinson's disease. *Journal of movement disorders*, 11(3), 121.

12. Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., ... & Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, 32, 195-214.

13. Meilán, J. J., Martínez-Sánchez, F., Carro, J., Sánchez, J. A., & Pérez, E. (2012). Acoustic markers associated with impairment in language processing in Alzheimer's disease. *The Spanish journal of psychology*, 15(2), 487-494.

14. Neubauer, C., Chollet, M., Mozgai, S., Dennison, M., Khooshabeh, P., & Scherer, S. (2017, November). The relationship between task-induced stress, vocal changes, and physiological state during a dyadic team task. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (pp. 426-432).

15. Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. *Computers in Human Behavior*, 64, 366-382.

16. Narendra, N. P., & Alku, P. (2018, September). Dysarthric Speech Classification Using Glottal Features Computed from Non-words, Words and Sentences. In *Interspeech* (pp. 3403-3407).

17. Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, 26(5), 388-397.

18. Schucker, B., & Jacobs Jr, D. R. (1977). Assessment of behavioral risk for coronary disease by voice characteristics. *Psychosomatic Medicine*, 39(4), 219-228.

19. Schuller, B., Steidl, S., Batliner, A., Hirschberg, J., Burgoon, J.K., Baird, A., Elkins, A., Zhang, Y., Coutinho, E., Evanini, K. (2016) The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language. Proc. *Interspeech* 2016, 2001-2005, doi: 10.21437/Interspeech.2016-129.

20. Scherer, S., Hammal, Z., Yang, Y., Morency, L. P., & Cohn, J. F. (2014). Dyadic Behavior Analysis in Depression Severity Assessment Interviews. *Proceedings of the ... ACM International Conference on Multimodal Interaction. ICMI (Conference), 2014*, 112–119. https://doi.org/10.1145/2663204.2663238.

21. Stel, M., Dijk, E. van, Smith, P. K., Dijk, W. W. van, & Djalal, F. M. (2012). Lowering the Pitch of Your Voice Makes You Feel More Powerful and Think More Abstractly. *Social Psychological and Personality Science*, 3(4), 497–502. https://doi.org/10.1177/1948550611427610.

22. Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., Nishimura, M., & Arai, T. (2018). Major depressive disorder discrimination using vocal acoustic features. *Journal of affective disorders*, 225, 214–220. https://doi.org/10.1016/j.jad.2017.08.038.

23. Wagner, J., Triantafyllopoulos, A., Wierstorf, H., Schmitt, M., Burkhardt, F., Eyben, F., & Schuller, B. W. (2023). Dawn of the transformer era in speech emotion recognition: closing the valence gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*

24. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems 6(2), 156-166 (2005). doi: 10.1109/TITS.2005.848368.

25. Zontone, P., Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., Formaggia, F., Minen, D., Minen, M., Savorgnan, C.: Car Driver's Sympathetic Reaction Detection Through Electrodermal Activity and Electrocardiogram Measurements". IEEE Transactions on Biomedical engineering, vol. 49(12), 3413-3424 (2020). doi: 10.1109/TBME.2020.2987168.

26. Lemonnier, S., Bremond, R., Baccino, T.: Gaze behavior when approaching an intersection: Dwell time distribution and comparison with a quantitative prediction. Transportation Research Part F: Traffic Psychology and Behaviour 35, 60-74 (2015). doi: 10.1016/j.trf.2015.10.015.

27. Jabon, M., Bailenson, J., Pontikakis, E., Takayama, L., Nass, C.: Facial expression analysis for predicting unsafe driving behavior. IEEE Pervasive Comput 10(4), 84-95 (2011). doi: 10.1109/MPRV.2010.46.

28. Mehler B., Reimer, B., Coughlin, J.F., Dusek, J.A.: Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transp Res Rec 2138(1), 6-12 (2009). doi: 10.3141/2138-02.

29. Collet, C., Clarion, A., Morel, M., Chapon, A., Petit, C.: Physiological and behavioral changes associated to the management of secondary tasks while driving. App Ergon 40(6), 1041-1046 (2009).

30. Wierwille, W.W., Eggemeier, F.T.: Recommendations for mental workload measurement in a test and evaluation environment. Hum Factors 35(2), 263-281 (1992). doi: 10.1177/001872089303500205.

31. Evans, A., Boonstra, F. M. C., Perera, T., Vogel, A. P., Kolbe, S. C., Butzkueven, H., … van der Walt, A. (2018). What speech can tell us: A systematic review of dysarthria characteristics in Multiple Sclerosis. Autoimmunity Reviews, 17(12), 1202–1209. https://doi.org/10.1016/j.autrev.2018.06.010.

32. Skodda, S., Grönheit, W., Lukas, C., Bellenberg, B., Von Hein, S. M., Hoffmann, R., & Saft, C. (2016). Two different phenomena in basic motor speech performance in premanifest Huntington disease. Neurology, 86(14), 1329–1335. https://doi.org/10.1212/WNL.0000000000002550.

33. Martínez-Sánchez, F., Meilán, J. J. G., Carro, J., & Ivanova, O. (2018). A Prototype for the Voice Analysis Diagnosis of Alzheimer's Disease. Journal of Alzheimer's Disease, 64(2), 473–481. https://doi.org/10.3233/JAD-180037.
34. Schuller, B., Friedmann, F., & Eyben, F. (2013). Automatic recognition of physiological parameters in the human voice: Heart rate and skin conductance. In ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings (pp. 7219–7223). https://doi.org/10.1109/ICASSP.2013.6639064.
35. Koyuncu, E., Çam, P., Altınok, N., Çallı, D. E., Duman, T. Y., & Özgirgin, N. (2016). Speech and language therapy for aphasia following subacute stroke. Neural Regeneration Research, 11(10), 1591–1594. https://doi.org/10.4103/1673-5374.193237.
36. Moore, R. K.: Appropriate voices for artefacts: some key insights. In 1st Int. Workshop on Vocal Interactivity in-andbetween Humans, Animals and Robots (VIHAR-2017). 2017.
37. Mori, M.: The Uncanny Valley. Energy, 7(4), pp. 33–35, 1970.
38. Wilson, S. and Moore, R. K.: Robot, alien and cartoon voices: implications for speech-enabled systems In 1st Int.
Workshop on Vocal Interactivity in-and-between Hu- mans, Animals and Robots (VIHAR-2017). 2017.
39. Nass, C. and Brave, S.: Wired for Speech. MIT Press 2005.
40. C. Saint-Georges, M. Chetouani, R. Cassel, F. Apicella, A. Mahdhaoui, F. Muratori, M.-C. Laznik, and D. Cohen, "Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review)," PLoS ONE, vol. 8, no. 10, pp. 1–17, 2013. 593.
41. M. Spinelli, M. Fasolo, and J. Mesman, "Does prosody make the difference? a meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes," Developmental Review, vol. 44, pp. 1–18, 2017.
42. C. Kitamura and D. Burnham, "Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year." Infancy, vol. 4, pp. 85–110, 2003.
43. D. Burnham, U. Vollmer-Conna, and C. Kitamura, "What's new, Pussicat? on talking to babies and animals," Science, vol. 296, p. 1435, 2002.
44. L. J. Trainor, C. M. Austin, and R. N. Desjardins, "Is infantdirected speech prosody a result of the vocal expression of emotion?" Psychological Science, vol. 11, no. 3, 2000.
45. C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, The measurement of meaning. University of Illinois press, 1957, no. 47.

46. K. Scherer, R. Banse, H. Wallbott, and T. Goldbeck, “Vocal cues in emotion encoding and decoding,” Motivation and emotion , vol. 15, no. 2, pp. 123–148, 1991.