

Look Once to Hear



• شي مرات ف البلايص لي مزدحمة، بحال حفلة ولا شارع عامر، كيكون صعب على المخ البشري يركز ويفهم الهضرة ديال شخص معين اللي كتكون كتهضر معاه، خصوصا إلا كانو ناس آخرين كيهضرو حده. هاد السيستم الجديد الذكي ديال السماعات كيعالج هاد المشكلة بحيث كيتمكن ليك تعزل و تحسن الصوت ديال الشخص اللي باغي تسمع له، و تصفي جميع الأصوات الأخرى و الضجيج اللي ف الخلفية.

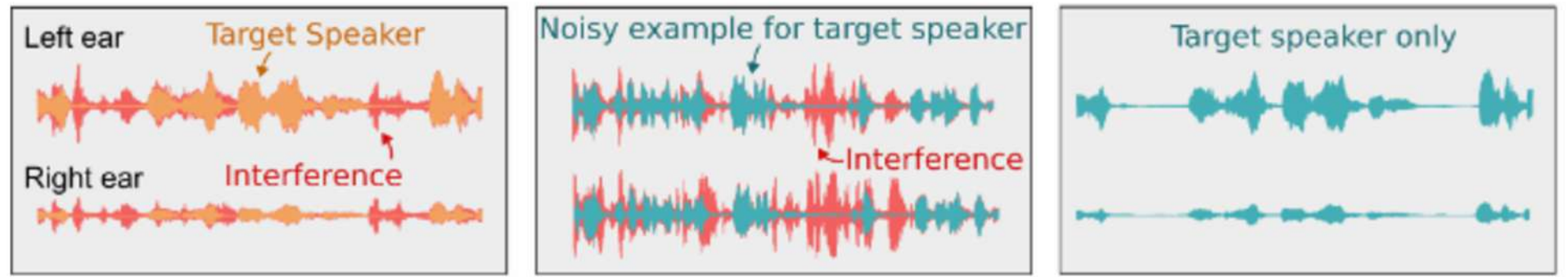
• المفتاح ديال هاد الاختراع هو أن السيستم كييفي غير عينة قصيرة و مشوشة من الصوت ديال المتكلم المستهدف باش ييدا يخدم. كيکفي غير تركز على الشخص اللي بغيت تركز معاه لبضع ثواني كيکفا تقدر تشوفو فالتصاور، و السيستم كيلتقط هداك العينة القصيرة ديال الصوت. من بعد، كيستعمل هديك العينة باش يتعلم الخصائص ديال الصوت ديال هداك الشخص و يقدر من بعد يعزل و يعزز الهضرة ديالو، حتى ف وسط الجماعة.

• هاد الشي مريح بزاف مقارنة مع الطرق القديمة اللي كانو كيتطلبو تسجيل نقي وبجودة عالية لصوت المتكلم لي باغي تسمع له. الحصول على هاد العينة النقية كيكون غالبا صعب ف البيئات الحقيقية و الواقع، اللي غالبا فيها الضجيج. باستخدام عينة سريعة و غير معدلة، هاد السيستم الجديد كييقي أسهل ف الاستعمال ف المواقف العملية و قادر يعالج الصوت ف real time.

• ولكن، واحد الاشكال فهاد paper لي دار هاد الخدمة هو ما كيتعمقش ف الاحتمالات المحدودة أو المجالات اللي كتححتاج مزيد من البحث. مثلا، غيكون مثير للاهتمام مثلا نفهمو كيفاش السيستم كيخدم مع جماعة لي كيهضرو ف وقت واحد، أولا كيهضرو لهجات ولغات مختلفة، وكذا القوة ديالو ف مواجهة أنواع ومستويات مختلفة ديال الضجيج والتشويش.

References:

Look Once to Hear: Target Speech Hearing with Noisy Examples. <https://arxiv.org/pdf/2405.06289>



• هاد البحث كييقدّم نظام ذكي جديد للسماعات اللي كيتمكن بفعالية يعزل ويحسن صوت المتكلم المستهدف ف وجود هضرة وضجيج ف الخلفية. الشيء لي جديد و نقدرو نعتابروه state-of-the-art ديال الابتكار هو قدرة السيستم على الخدمة بعينة صوت قصيرة و حقيقية بدون تغيير ديال المتكلم لي مستهدف، اللي كتخاد منين حامل السماعات كيركز عليه لبضع ثواني، كيکفا تقدر تشوفو فالتصاور.

• ورغم أن هاد البحث ما كيطرش بزاف نتاع الامور، إلا ان هاد الدراسة كتتمثل خطوة مهمة لقدام ف مجال معالجة الصوت الذكي وتطبيقه لتحسين التجارب البشرية.

Claude 3.5 Sonnet: Leading the Way in AI Task Management

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro
Visual math reasoning <i>MathVista (testmini)</i>	67.7% 0-shot CoT	50.5% 0-shot CoT	63.8% 0-shot CoT	63.9% 0-shot CoT
Science diagrams <i>AIZD, test</i>	94.7% 0-shot	88.1% 0-shot	94.2% 0-shot	94.4% 0-shot
Visual question answering <i>MMMU (val)</i>	68.3% 0-shot CoT	59.4% 0-shot CoT	69.1% 0-shot CoT	62.2% 0-shot CoT
Chart Q&A <i>Relaxed accuracy (test)</i>	90.8% 0-shot CoT	80.8% 0-shot CoT	85.7% 0-shot CoT	87.2% 0-shot CoT
Document visual Q&A <i>ANLS score, test</i>	95.2% 0-shot	89.3% 0-shot	92.8% 0-shot	93.1% 0-shot

• طورو الباحثين ف Anthropic — فريق مخصص لخلق نماذج نقدر نوجهوها و نتيقو فيها— موديل Claude 3.5 Sonnet كيحط معايير جديدة للتفكير فمستوى الدراسات العليا (GPQA)، المعرفة فمستوى البكالوريا (MMLU)، وإتقان البرمجة (HumanEval). الموديل كيبيين تحسن بزاف ففهم التفاصيل الصغيرة، والفكاهة، والتعليمات المعقدة و قادر ينتج محتوى عالي الجودة بنبرة طبيعية ومفهومة.

• Claude 3.5 Sonnet كيشغل بسرعة ضعف اللي كيشغل بيه Claude 3 Opus، وهاد الزيادة فالأداء مع الثمن المناسب كيخليه مثالي لمصالح دعم الزبناء و تدبير workflows لي فيهم خطوات المتعددة.

• فالتقييم الداخلي للبرمجة، Claude 3.5 Sonnet كيقدر يصلح الأخطاء ولا يزيد خصائص لقاعدة بيانات مفتوحة المصدر من خلال وصف بالكتابة، وكيكتب، ويعدل، ويشغل الكود بوحده بفهم معالجة الترجمة ديال الكود بكفاءة، وهادشي كيخليه فعال لتحديث التطبيقات القديمة ونقل قواعد البيانات.

• كأقوى نموذج رؤية حتى لدبا من Anthropic، كيتفوق فمهام الاستدلال البصري بحال تفسير الرسوم البيانية والجداول، وكيقدر بدقة ينقل النصوص من الصور الناقصة، وهادشي كييفيد ناس د الكوميرس، اللوجستيك، والخدمات المالية بتفكيك الفكرة ديال الصور، الرسوم البيانية، أو التوضيحات بدل ما تكون بالكتابة بوحدها.

References:

- The Claude 3 Model Family: Opus, Sonnet, Haiku: <https://www-cdn.anthropic.com/f2986af8d052f26236f6251da62d16172cfabd6e/claude-3-model-card.pdf>

official partner



www.mathemaroc.com

math&maroc



اسمع من الوهلة الأولى : Look Once to Hear



• بعض الأحيان في الأماكن المزدحمة، مثل الحفلات أو الشوارع المليئة بالناس، يكون من الصعب على الدماغ البشري التركيز وفهم كلام شخص معين يتحدث معه، خاصة إذا كان هناك أشخاص آخرون يتحدثون بالقرب منه. هذا النظام الذكي الجديد للسماعات (Smart Audio System) يعالج هذه المشكلة من خلال تمكينك من عزل وتحسين صوت الشخص الذي تريد الاستماع إليه، وتنقية جميع الأصوات الأخرى والضجيج الخلفي (Background Noise).

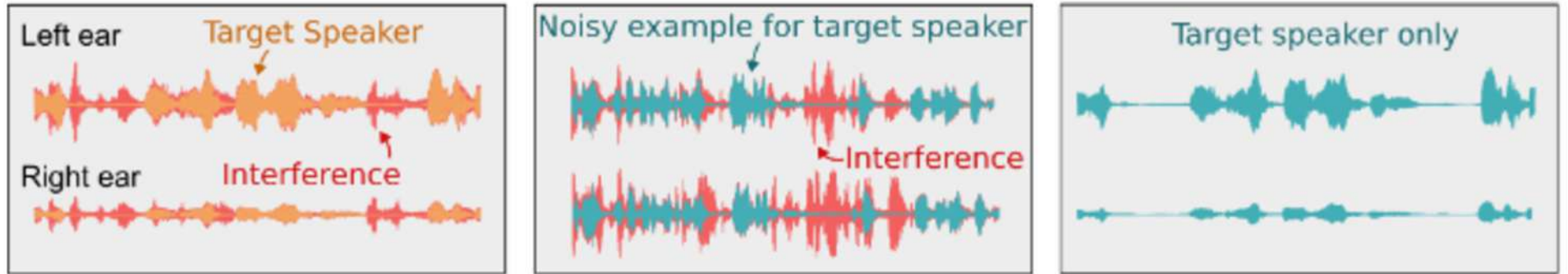
• المفتاح لهذا الاختراع هو أن النظام يحتاج فقط إلى عينة قصيرة ومشوشة من صوت المتحدث المستهدف ليبدأ في العمل. يكفي التركيز على الشخص المراد الاستماع إليه لبضع ثوانٍ، كما يظهر في الصور المرفقة. يقوم النظام بالتقاط هذه العينة القصيرة من الصوت. بعد ذلك، يستخدم النظام هذه العينة لتعلم خصائص صوت ذلك الشخص، مما يمكنه من عزل وتعزيز الصوت المستهدف حتى في البيئات المزدحمة.

• هذا النظام يوفر مزايا عديدة مقارنة بالطرق التقليدية التي كانت تتطلب تسجيل صوت المتحدث بجودة عالية. فالحصول على عينة صوت نقية كان غالبًا ما يكون صعبًا في البيئات الحقيقية التي تكون فيها الضوضاء متواجدة بشكل كبير. باستخدام عينة قصيرة وغير معدلة، يبقى هذا النظام أسهل في الاستخدام في السيناريوهات العملية، حيث يستطيع معالجة الصوت في الوقت الحقيقي (Realtime) بفعالية وبدقة.

• في بعض الأوراق البحثية التي تناولت هذه الخدمة، يظهر أحد العيوب هو عدم التعمق في الاحتمالات المحدودة أو المجالات التي تتطلب دراسة إضافية. على سبيل المثال، من المثير للاهتمام فهم كيفية تشغيل النظام مع مجموعات تتحدث في نفس الوقت، مع استخدام لهجات ولغات متنوعة، وكذلك قدرته على التعامل مع مستويات متفاوتة من الضوضاء والتشويش.

المراجع:

Look Once to Hear: Target Speech Hearing with Noisy Examples.
<https://arxiv.org/pdf/2405.06289>



• هذا البحث يقدم نظامًا ذكيًا جديدًا للسماعات الذي يمكنه بفعالية عزل وتحسين صوت المتحدث المستهدف في وجود هضمة وضجيج في الخلفية. ما يميز هذا الابتكار ويجعله يستحق الاعتبار كحالة من حالات الفن هو قدرة النظام على العمل بعينة صوت قصيرة وحقيقية دون تغيير في صوت المتحدث المستهدف، حيث يتم التركيز عليه لبضع ثوانٍ كما هو موضح في الصور.

• على الرغم من أن هذا البحث لا يغطي جميع جوانب الموضوع، إلا أنه يمثل خطوة هامة للأمام في مجال معالجة الصوت الذكي وتطبيقه لتحسين تجارب المستخدمين.

الرائد في إدارة مهام الذكاء الاصطناعي : Claude 3.5 Sonnet

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro
Visual math reasoning <i>MathVista (testmini)</i>	67.7% 0-shot CoT	50.5% 0-shot CoT	63.8% 0-shot CoT	63.9% 0-shot CoT
Science diagrams <i>AI2D, test</i>	94.7% 0-shot	88.1% 0-shot	94.2% 0-shot	94.4% 0-shot
Visual question answering <i>MMMU (val)</i>	68.3% 0-shot CoT	59.4% 0-shot CoT	69.1% 0-shot CoT	62.2% 0-shot CoT
Chart Q&A <i>Relaxed accuracy (test)</i>	90.8% 0-shot CoT	80.8% 0-shot CoT	85.7% 0-shot CoT	87.2% 0-shot CoT
Document visual Q&A <i>ANLS score, test</i>	95.2% 0-shot	89.3% 0-shot	92.8% 0-shot	93.1% 0-shot

• تم تطوير Claude 3.5 Sonnet بواسطة الباحثين في Anthropic—فريق مكرس لإنشاء نماذج أكثر أمانًا وقابلية للتوجيه وموثوقية—ويضع معايير جديدة للتفكير على مستوى الدراسات العليا (GPQA)، والمعرفة على مستوى البكالوريوس (MMLU)، وإتقان البرمجة (HumanEval). يُظهر تحسُّنًا ملحوظًا في فهم الفروق الدقيقة والفكاهة والتعليمات المعقدة، ويتفوق في كتابة المحتوى عالي الجودة بنبرة طبيعية وجذابة.

• يعمل Claude 3.5 Sonnet بسرعتين ضعف سرعة Claude 3 Opus، مما يجمع بين تعزيز الأداء وتسعير فعال من حيث التكلفة، مما يجعله مثاليًا لدعم العملاء الحساسين للسياق وإدارة تدفقات العمل متعددة الخطوات.

• في تقييم الترميز الداخلي، يمكن لـ Claude 3.5 Sonnet إصلاح الأخطاء أو إضافة وظائف إلى قاعدة بيانات open source من وصف طبيعي للغة، وكتابة وتحرير وتنفيذ الكود بشكل مستقل، ومعالجة ترجمات الكود بكفاءة، مما يجعله فعالًا في تحديث التطبيقات القديمة ونقل قواعد البيانات.

• باعتباره أقوى نموذج رؤية حتى الآن من Anthropic، يتفوق في مهام الاستدلال البصري مثل تفسير الرسوم البيانية والجداول، ويمكنه نسخ النصوص بدقة من الصور غير الكاملة، مما يفيد البيع بالتجزئة والخدمات اللوجستية والخدمات المالية من خلال استخراج المزيد من الأفكار من الصور أو الرسوم البيانية أو الرسوم التوضيحية أكثر من النصوص وحدها.

مراجع:

The Claude 3 Model Family: Opus, Sonnet, Haiku: <https://www-cdn.anthropic.com/f2986af8d052f26236f6251da62d16172cfabd6e/claude-3-model-card.pdf>

official partner



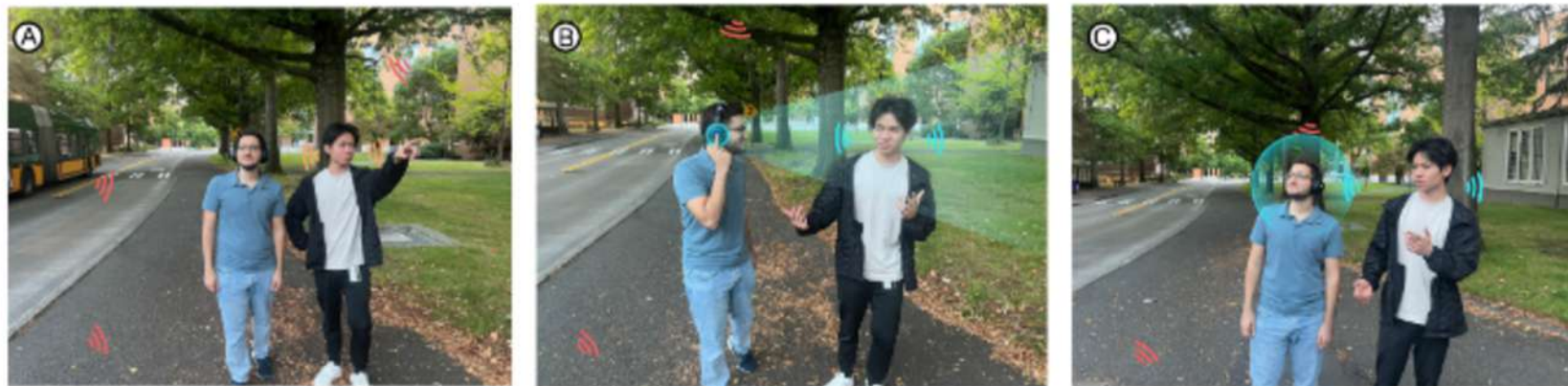
www.mathemaroc.com

math&maroc

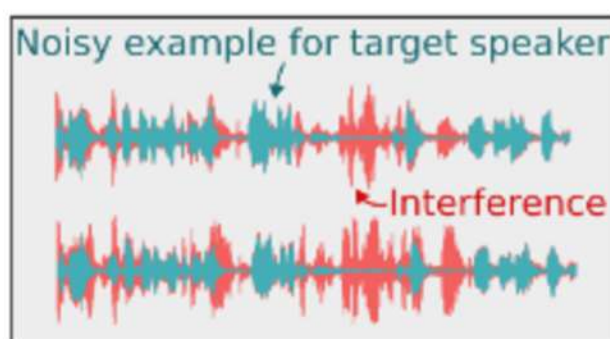
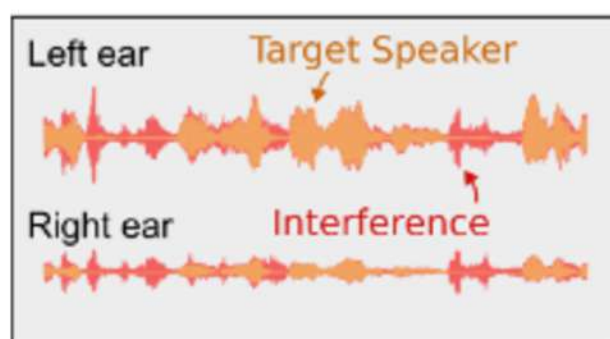


Look Once to Hear

- Sometimes in crowded places, like parties or busy streets, it can be difficult for the human brain to focus and understand the speech of a specific person talking, especially with others nearby. This new Smart Audio System addresses this issue by allowing you to isolate and enhance the voice of the person you want to listen to, filtering out all other voices and background noise.
- The key to this invention is that the system only needs a short and noisy sample of the speaker's voice to start working. Simply focus on the person you want to listen to for a few seconds, as shown in the attached images. The system captures this short voice sample, then uses this sample to learn the characteristics of that person's voice, enabling it to isolate and enhance their voice even in noisy surroundings.



- This system grants several advantages over traditional methods that required high-quality recordings of the speaker's voice. Obtaining a pure voice sample was often deemed difficult in real-world environments where noise is prevalent. Using a short and unmodified sample, this system remains easier to use in practical scenarios, capable of processing sound in real time effectively and accurately.
- In some research papers discussing the matter, one drawback highlighted is the lack of depth in exploring limited possibilities or areas requiring further study. For instance, it would be interesting to understand how the system operates with groups speaking simultaneously, using different accents and languages, as well as its ability to handle varying levels of noise and interference.



References:

Look Once to Hear: Target Speech Hearing with Noisy Examples. <https://arxiv.org/pdf/2405.06289>

- This research introduces a new intelligent audio system that effectively isolates and enhances the voice of the targeted speaker amidst background digestion and noise. What sets this innovation apart and makes it noteworthy as state-of-art is its ability to operate with a short and real voice sample without altering the voice of the targeted speaker, focusing on them for a few seconds as illustrated in the images.
- While this research does not encompass every facet of the subject, it represents an important step forward in the field of intelligent sound processing and its application to enhance user experiences.

Claude 3.5 Sonnet: Leading the Way in AI Task Management

- Claude 3.5 Sonnet, developed by researchers at Anthropic—a team dedicated to creating safer, steerable, and more reliable models—sets new benchmarks for graduate-level reasoning (GPQA), undergraduate knowledge (MMLU), and coding proficiency (HumanEval). It shows marked improvement in grasping nuance, humor, and complex instructions, and excels at writing high-quality content with a natural, relatable tone.

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro
Visual math reasoning <i>MathVista (testmini)</i>	67.7% 0-shot CoT	50.5% 0-shot CoT	63.8% 0-shot CoT	63.9% 0-shot CoT
Science diagrams <i>AI2D, test</i>	94.7% 0-shot	88.1% 0-shot	94.2% 0-shot	94.4% 0-shot
Visual question answering <i>MMMU (val)</i>	68.3% 0-shot CoT	59.4% 0-shot CoT	69.1% 0-shot CoT	62.2% 0-shot CoT
Chart Q&A <i>Relaxed accuracy (test)</i>	90.8% 0-shot CoT	80.8% 0-shot CoT	85.7% 0-shot CoT	87.2% 0-shot CoT
Document visual Q&A <i>ANLS score, test</i>	95.2% 0-shot	89.3% 0-shot	92.8% 0-shot	93.1% 0-shot

- Operating at twice the speed of Claude 3 Opus, Claude 3.5 Sonnet combines performance boost with cost-effective pricing, making it ideal for context-sensitive customer support and multi-step workflows.
- In internal coding evaluation, Claude 3.5 Sonnet can fix bugs or add functionality to an open source codebase from a natural language description, independently write, edit, and execute code, and handle code translations efficiently, making it effective for updating legacy applications and migrating codebases.
- As the strongest vision model yet by Anthropic, it excels in visual reasoning tasks like interpreting charts and graphs and can accurately transcribe text from imperfect images, benefiting retail, logistics, and financial services by extracting more insights from images, graphics, or illustrations than from text alone.

References:

- The Claude 3 Model Family: Opus, Sonnet, Haiku: <https://www-cdn.anthropic.com/f2986af8d052f26236f6251da62d16172cfabd6e/claude-3-model-card.pdf>

official partner



www.mathemaroc.com



math&maroc