شرح المشروع بالعربية

تحليل إشارة الصوت باستخدام تعلّم الآلة للكشف المبكر عن مرض باركنسون

دليل دراسي من تسعة عشر جزءاً

من المبادئ الأولى إلى تحضير المناقشة

سهيل محمد الحمالي

رقم القيد 2200208522

المشرف: الدكتور عادل أضياف

قسم الهندسة الطبية الحيوية، كلية الهندسة، جامعة طرابلس

كل رقم في هذا الدليل مأخوذ من مجلد `docs/` ومجلد `reports/` في المشروع.

المحتويات

الجزء 0 --- خريطة الطريق

أولاً: ما هو مشروعك في جملة واحدة

مشروعك يجيب على سؤال واحد محدّد:

«هل نستطيع أن نقيس تسجيلاً صوتياً لشخص ما، ونقول إن نمطه الصوتي يشبه أصوات المصابين بمرض باركنسون أم يشبه أصوات الأشخاص الأصحّاء، باستخدام تعلّم آلة كلاسيكي قابل للتفسير فقط؟»

لاحظ الصياغة بدقّة: السؤال هو «**يشبه أي مجموعة**»، وليس «هل هذا الشخص مريض». هذا الفرق ليس تجميلاً لغوياً، بل هو جوهر المشروع علمياً وأخلاقياً، وسنعود إليه كثيراً.

المخرج النهائي هو **أداة دعم فرز بحثية غير تشخيصية** (research screening-support prototype): برنامج صوتياً، فيعطيك مؤشراً حذراً غير طبّي، مع كامل الأدلة العلمية على مدى جودته وأين يفشل. المصدر: *docs/01_overview_and_journey.md* القسم *1.1*

ثانياً: لماذا الصوت أصلاً؟

مرض باركنسون يقلّل مادة الدوبامين في مناطق الدماغ المسؤولة عن الحركة. وأعراضه المعروفة، وهي الرعاش والتيبّس وبطء الحركة، لا تصيب اليدين والقدمين فقط، بل تصيب **كل عضلة**، بما فيها العضلات الصغيرة جداً التي تتحكّم بالحبال الصوتية واللسان والشفتين والتنفّس.

ولأن هذه العضلات صغيرة ودقيقة جداً، فإن تأثّرها قد يظهر **مبكّراً**، أحياناً قبل أن تصبح أعراض الحركة واضحة بما يكفي لدفع الشخص لزيارة الطبيب. وهنا تكمن الفكرة كلّها: الميكروفون رخيص، وغير جارح، ولا يحتاج مستشفى. المصدر: *docs/02_scientific_background.md* القسم *2.2*

ثالثاً: الأرقام التي يجب أن تحفظها

هذه هي أهم أرقام المشروع، وكلّها من docs/README.md:

- عدد الأشخاص في بيانات التدريب: 37 شخصاً (21 سليماً و16 مريضاً).
- عدد التسجيلات الصوتية: 73 تسجيلاً.
- عدد المقاطع بعد التقطيع إلى 10 ثوانٍ: 1001 مقطع.

- عدد الخصائص الصوتية المستخرجة من كل مقطع: 74 خاصية.
- عدد التكوينات النموذجية التي جُرّبت وقورنت: 19 تكويناً.
- الدقة المتوازنة داخلياً على مستوى الشخص: 0.822 مع انحراف معياري 0.032.
- مساحة تحت المنحنى (ROC-AUC) داخلياً: 0.864.
- الدقة المتوازنة خارجياً على البيانات الإيطالية: 0.629.
- مساحة تحت المنحنى خارجياً: 0.701.
- النتيجة عند استخدام تقسيم خاطئ متعمَّد: 0.909، وهي **رقم مضخَّم وليس نتيجة**.

لا تقلق إن لم تفهم معنى «الدقة المتوازنة» أو «المساحة تحت المنحنى» الآن، فالجزء 10 كامل مخصّص لشرحها بأمثلة رقمية بسيطة.

رابعاً: النقطة التي تميّز مشروعك

هناك أمر يجب أن تفهمه من البداية لأنه سيتكرّر في كل جزء تقريباً، وهو أهم ما ستدافع عنه أمام الأساتذة.

كثير من الأبحاث المنشورة على هذه النوعية من البيانات تُعلن نتائج بين 95٪ و99٪. ومشروعك يُعلن 82٪. للوهلة الأولى يبدو هذا ضعفاً. لكنه في الحقيقة **العكس تماماً**.

السبب أن تلك الأبحاث غالباً تخلط تسجيلات الشخص نفسه بين بيانات التدريب وبيانات الاختبار. فيصبح النموذج قادراً على «التعرّف على الشخص» بدل «التعرّف على المرض»، تماماً كطالب رأى أسئلة الامتحان قبل الامتحان.

وأنتم لم تكتفوا بقول ذلك، بل **قِستموه**: نفس الأنبوب البرمجي بالضبط، ونفس النموذج بالضبط، يعطي 0.909 بالتقسيم الخاطئ و0.775 بالتقسيم الصحيح. الفارق ليس قدرة على الكشف، بل حفظ. المصدر: `docs/05_validation_and_metrics.md` القسم *5.4*

خامساً: أربع نقاط قوة ستقولها في المناقشة

- بنيتم **القياس** نفسه، لا النموذج فقط. استخراج 74 خاصية صوتية من الصوت الخام هو المساهمة الهندسية الأساسية، ولم تُحمَّل جاهزة من أي مكان.
- تحقّقتم **بأمانة**: لا يظهر أي شخص في التدريب والاختبار معاً، وهذا مفروض بسطر في الكود يُوقف البرنامج فوراً إن حدث.
- **قِستم** قدرة التعميم بدل افتراضها: اختبرتم النموذج مجمّداً على بيانات إيطالية بلغة أخرى، فأعطى 0.701. قلّة قليلة من مشاريع البكالوريوس تفعل هذا.
- بنيتم نظاماً **يعترف بعدم اليقين**: عندما تكون الدرجة قريبة من الحدّ الفاصل يقول «غير حاسم» بدل أن يعطي تصنيفاً واثقاً.

المصدر: `docs/12_presentation_qa.md` القسم *12.1*

سادساً: مسار المشروع الزمني باختصار

لكي تتخيّل الصورة الكبيرة، هذا ما حدث بالترتيب:

أولاً بُنيت بنية المشروع وبيئة العمل، ثم فُحصت البيانات فحصاً كاملاً **قبل** أي تدريب، ثم بُني أنبوب المعالجة واستخراج الخصائص، ثم دُرّبت النماذج وقيّمت تقييماً نزيهاً فأعطت 0.780، ثم بُني التطبيق، ثم طلب المشرف تحسين الدقة فأُجريت دراسة من 19 تكويناً رفعت النتيجة إلى 0.822، ثم سجّل المشرف صوته وصوت صديقه، وكلاهما سليم، فحصل كلاهما على درجة قريبة من 0.65. وهذه الحادثة كشفت ضعفاً حقيقياً وأدّت إلى إضافة «النطاق غير الحاسم». وأخيراً جاء التحقّق الخارجي على البيانات الإيطالية. المصدر: `docs/01_overview_and_journey.md` القسم *1.3*

لاحظ شيئاً مهماً هنا: أهم تحسينات المشروع لم تأتِ من نجاح، بل من **ملاحظة فشل**. هذه قصة جيدة جداً لتحكيها في المناقشة.

خلاصة الجزء

- المشروع يقيس تشابه نمط صوتي مع مجموعة بيانات، ولا يشخّص مرضاً إطلاقاً.
- الأرقام الأساسية: 37 شخصاً، و73 تسجيلاً، و74 خاصية، ونتيجة 0.822 داخلياً و0.701 خارجياً.
- قوة المشروع ليست في ارتفاع الرقم، بل في نزاهة طريقة قياسه، والدليل المقيس على ذلك هو الفرق بين 0.909 و0.775.

الجزء 1 --- المشكلة الطبية وكيف يُنتَج الصوت

قبل أن نتحدث عن أي كود أو أي نموذج، يجب أن تفهم شيئاً واحداً فهماً عميقاً: **لماذا يمكن للصوت أصلاً أن يحمل معلومة عن مرض في الدماغ؟**

إن فهمت هذا الجزء جيداً، فسيصبح كل ما بعده منطقياً وطبيعياً. وإن لم تفهمه، فسيبدو المشروع كأنه أرقام عشوائية.

أولاً: الصوت البشري يُنتَج على ثلاث مراحل

الكلام ليس عملية واحدة، بل ثلاث مراحل متتابعة. والتشبيه الأدقّ هو **آلة موسيقية نفخية**، مثل المزمار.

المرحلة الأولى: مصدر الطاقة (the power source) --- الرئتان. الرئتان (lungs) تدفعان الهواء إلى أعلى عبر القصبة الهوائية (trachea). هذا الهواء المتدفّق هو الطاقة التي ستتحوّل لاحقاً إلى صوت. لا هواء يعني لا صوت إطلاقاً. في تشبيه الآلة الموسيقية: الرئتان هما **الشخص الذي ينفخ**.

المرحلة الثانية: المُهتزّ (the vibrator) --- الحبال الصوتية. في الحنجرة (larynx) توجد طيّتان عضليّتان صغيرتان تُسمّيان الحبال الصوتية (vocal folds). عندما تقرّبهما من بعضهما، يصطدم الهواء الصاعد بهما فيفتحهما بالقوّة، ثم تعود مرونتهما الطبيعية فتُغلقهما بسرعة، ثم يفتحهما الهواء مرة أخرى، وهكذا **مئات المرات في الثانية الواحدة**. كل دورة فتح وإغلاق تُطلق «نفخة» صغيرة من الهواء، وتتابع هذه النفخات السريعة هو الصوت الخام الطنّان للإنسان. في تشبيه الآلة: الحبال الصوتية هي **ريشة المزمار**.

المرحلة الثالثة: المرشِّح (the filter) --- مجرى الصوت. الصوت الخام الخارج من الحنجرة يمرّ بعد ذلك في البلعوم والفم واللسان والشفتين، وهذه كلها معاً تُسمّى مجرى الصوت (vocal tract). هذه المنطقة تعمل كأنبوب رنين: تحريك اللسان والشفتين يُقوّي بعض الترددات ويُضعف أخرى، فيتحوّل الطنين الخام إلى حروف متحرّكة (vowels) وحروف ساكنة (consonants). في تشبيه الآلة: مجرى الصوت هو **جسم الآلة وفتحاتها**. المصدر: *docs/02_scientific_background.md* القسم *2.1*

ثانياً: نتيجتان مهمّتان جداً لمشروعك

النتيجة الأولى: سرعة اهتزاز الحبال الصوتية هي «طبقة الصوت». عدد مرات فتح وإغلاق الحبال الصوتية في الثانية يُسمّى التردد الأساسي (fundamental frequency) ويُرمز له بـF0، ويُقاس بالهرتز (Hz). وأنت تسمعه كـ«حدّة» أو «غِلَظ» الصوت.

- الرجل البالغ: تقريباً بين 85 و180 هرتز.

- المرأة البالغة: تقريباً بين 165 و255 هرتز.

انتبه جيداً لهذه النقطة، فهي مصدر مشكلة سنواجهها لاحقاً: طبقة الصوت المطلقة تحمل معلومة عن **جنس المتحدّث**، لا عن مرضه. وهذا ما يُسمّى «عاملاً مُربكاً» (confounder)، أي متغيّر خفي مرتبط بالجواب ويسمح للنموذج بأن «يغشّ». وفي الجزء 11 سأشرح كيف اختبرتم هذه المشكلة بدل تجاهلها.

النتيجة الثانية: انتظام الاهتزاز هو «جودة الصوت». في الصوت السليم، الاهتزاز **دوري بدرجة عالية** (highly periodic): كل دورة تكاد تكون نسخة طبق الأصل من الدورة السابقة، نفس المدّة الزمنية تقريباً، ونفس القوة تقريباً. وهذه الجملة تحديداً هي الأساس العلمي لأهم خصائص مشروعك: jitter وshimmer وHNR. فكلّها تقيس شيئاً واحداً بصور مختلفة: **إلى أي درجة تشبه كل دورة الدورة التي قبلها؟**

ثالثاً: ماذا يفعل باركنسون بالضبط؟

مرض باركنسون (Parkinson's disease) مرض تنكّسي عصبي سببه نقص مادة الدوبامين (dopamine) في مناطق من الدماغ مسؤولة عن التحكّم بالحركة. وأعراضه الكلاسيكية أربعة:

- الرعاش (tremor): اهتزاز لا إرادي.
- التيبّس (rigidity): تصلّب العضلات.
- بطء الحركة (bradykinesia): الحركة تصير أبطأ.
- قِلّة سعة الحركة (hypokinesia): الحركة تصير **أصغر** في المدى.

والنقطة الجوهرية التي يجب أن تقولها في المناقشة: **هذه الأعراض تصيب كل عضلة في الجسم، لا اليدين فقط.** والعضلات التي تتحكّم بالحبال الصوتية واللسان والشفتين والتنفّس هي عضلات صغيرة جداً وتتطلّب تحكّماً دقيقاً جداً، ولهذا فهي من أوائل ما يتأثّر.

واضطراب الكلام الناتج عن هذا له اسم طبي محدّد: **عُسر التلفّظ ناقص الحركة** (hypokinetic dysarthria).

- عُسر التلفّظ (dysarthria): اضطراب في النطق سببه ضعف أو سوء تحكّم في عضلات الكلام.
- عُسر الصوت (dysphonia): اضطراب في جودة الصوت نفسه، كالبحّة والهَمْس.

رابعاً: الجسر بين العَرَض الطبي والرقم في الحاسوب

هذه أهم قائمة في الجزء كلّه. كل بند فيها يربط بين ثلاثة أشياء: ما يراه الطبيب، وما يحدث فيزيائياً، وما نستطيع نحن قياسه بالحاسوب.

- **انخفاض الصوت والكلام الرتيب.** السبب الفيزيائي: قِلّة سعة الحركة وضعف دعم التنفّس. ما نقيسه: **انخفاض التغيّر** في F0 وفي السعة، أي أن الانحراف المعياري لطبقة الصوت ومداها يصغران.
- **صوت مهموس أو مبحوح.** السبب: الحبال الصوتية لا تنغلق انغلاقاً كاملاً، فيتسرّب هواء على شكل هواء مضطرب. ما نقيسه: **ضجيج أكثر مقابل نغمة أقل**، أي انخفاض HNR وCPPS.

- **صوت غير مستقرّ.** السبب: الرعاش وعدم انتظام التحكّم العضلي. ما نقيسه: **زيادة عدم الانتظام بين دورة وأخرى**، أي ارتفاع jitter وshimmer.
- **حروف ساكنة غير دقيقة.** السبب: حركات اللسان والشفتين صارت أبطأ وأصغر. ما نقيسه: **شكل طيفي مختلف وأبطأ تغيّراً**، أي MFCC ومشتقّاتها delta-MFCC.
- **تردّد وتوقّفات في غير محلّها.** السبب: صعوبة في بدء الحركة وفي الاستمرار بها. ما نقيسه: **توقّفات أكثر وأطول**، أي خصائص التوقّفات.

المصدر: *docs/02_scientific_background.md* القسم *2.2*

إن سألك أستاذ في المناقشة «لماذا اخترتم هذه الخصائص تحديداً؟»، فجوابك هو هذه القائمة بالضبط. كل خاصية في مشروعك لها مبرّر فسيولوجي، ولم تُختَر لأنها متوفّرة في مكتبة برمجية.

خامساً: لماذا «الكشف المبكّر» تحديداً؟

التغيّرات الصوتية التي ذكرناها كثيراً ما تظهر **مبكّراً**، أحياناً قبل أن تصبح الأعراض الحركية واضحة بما يكفي لتدفع الشخص إلى زيارة طبيب أعصاب. والسبب منطقي: عضلات الحنجرة واللسان أصغر بكثير من عضلات الذراع، وتحتاج دقّة تحكّم أعلى بكثير. فالنقص البسيط في التحكّم الحركي قد لا يظهر في مشيتك، لكنه يظهر في اهتزاز يتكرّر 150 مرة في الثانية.

ولهذا فإن فكرة الفرز الصوتي (voice-based screening) جذّابة عملياً:

- الميكروفون رخيص جداً ومتوفّر في كل هاتف.
- الفحص غير جارح (non-invasive) تماماً، فلا إبَر ولا أشعّة ولا مستشفى.
- يمكن تكراره بسهولة وبتكلفة شبه معدومة.
- يمكن، نظرياً، أن يُشير إلى الأشخاص الذين يُستحسن أن يراجعوا طبيب أعصاب.

لاحظ الصياغة الأخيرة: «يُشير إلى من يُستحسن أن يراجع طبيباً»، وليس «يُشخّص». هذا هو التعريف الدقيق لكلمة **فرز** (screening): خطوة تسبق التشخيص وتُوجّه إليه، ولا تحلّ محلّه أبداً.

سادساً: ما لا يدّعيه مشروعك

يجب أن تكون هذه الجملة حاضرة في ذهنك دائماً: **المشروع لا يدّعي أن هذه القياسات تُشخّص مرض باركنسون.**

والسبب بسيط ومقنع جداً: هناك حالات كثيرة أخرى تُغيّر جودة الصوت بنفس الاتجاه تقريباً: الزكام والتهاب الحنجرة (laryngitis)، والتقدّم في العمر، والتدخين، والإرهاق وقلّة النوم، وأمراض عصبية أخرى غير باركنسون.

فإذا قاس النظام ارتفاعاً في jitter وانخفاضاً في HNR، فهو لا يستطيع أن يميّز إن كان السبب باركنسون أم زكاماً. كل ما يستطيع قوله هو: «**هذا النمط الصوتي يشبه نمط المجموعة المرضية في مجموعة البيانات التي تدرّبتُ عليها.**» المصدر: *docs/02_scientific_background.md* نهاية القسم *2.2*

خلاصة الجزء

- الصوت يُنتَج على ثلاث مراحل: الرئتان مصدر الطاقة، والحبال الصوتية هي المُهتزّ الذي يحدّد F0، ومجرى الصوت هو المرشِّح الذي يصنع الحروف.
- باركنسون يُضعف التحكّم في كل العضلات بما فيها عضلات الكلام الدقيقة، فينتج `hypokinetic dysarthria`، ولكل عَرَض فيه مقابل رقمي نقيسه.
- الفكرة كلّها فرز مبكّر رخيص وغير جارح، وليست تشخيصاً، لأن حالات كثيرة أخرى تُغيّر الصوت بنفس الطريقة.

الجزء 2 --- الصوت كإشارة رقمية

في الجزء السابق عرفنا كيف يُنتج الجسم الصوت. الآن السؤال التالي: **كيف يدخل هذا الصوت إلى الحاسوب ويصير أرقاماً؟**

أولاً: الصوت في الأصل موجة ضغط

عندما تهتزّ حبالك الصوتية، فهي تضغط الهواء المجاور لها ثم تُرخيه، بالتناوب وبسرعة كبيرة. وهذا التتابع من الضغط والتخلخل ينتشر في الهواء كموجة، وتُسمّى موجة ضغط (pressure wave). تخيّل أنك رميت حجراً في بركة ماء: تنتشر دوائر من الارتفاع والانخفاض. الصوت هو الشيء نفسه لكن في الهواء وبثلاثة أبعاد.

فالصوت إذن، في أصله، ظاهرة **مستمرة** (continuous): لا توجد لحظة لا قيمة فيها. وهذا يمثّل مشكلة للحاسوب، لأنه لا يستطيع تخزين عدد لا نهائي من القيم.

ثانياً: كيف يحوّل الميكروفون الموجة إلى كهرباء

داخل الميكروفون (microphone) غشاء رقيق جداً. وعندما تصطدم به موجة الضغط يهتزّ بنفس نمط الموجة، وهذا الاهتزاز الميكانيكي يُحوَّل إلى **جهد كهربائي متغيّر**. فالميكروفون في جوهره مترجم: يترجم «ضغط هواء» إلى «جهد كهربائي». والجهد الكهربائي لا يزال إشارة مستمرة، أي تناظرية (analogue).

ثالثاً: العيّنة، أهم مفهوم في هذا الجزء

هنا يأتي دور جهاز يُسمّى المحوّل التناظري الرقمي (analogue-to-digital converter، اختصاراً ADC)، وهو موجود في كل هاتف وكل حاسوب. وعمله بسيط جداً في الفكرة: **يقيس قيمة الجهد الكهربائي، ويسجّلها كرقم، ثم ينتظر جزءاً صغيراً جداً من الثانية، ثم يقيس مرة أخرى، ويكرّر ذلك آلاف المرات في الثانية**. وكل قياس منفرد من هذه القياسات يُسمّى **عيّنة** (sample). المصدر: *docs/11_glossary.md* قسم معالجة الصوت والإشارة

تشبيه: كاميرا الفيديو. الحركة في الواقع مستمرة، لكن الكاميرا لا تُسجّل «الحركة»، بل تلتقط **صوراً ثابتة متتابعة** بسرعة كبيرة، مثلاً 30 صورة في الثانية. وعندما تُعرَض بنفس السرعة، تخدع عينك فتراها حركة متصلة. العيّنة الصوتية هي «الصورة الثابتة»، والصوت الرقمي شريط من ملايين الصور الثابتة المتتابعة.

رابعاً: معدّل العيّنات

عدد العيّنات المأخوذة في الثانية الواحدة يُسمّى معدّل العيّنات (sample rate)، ويُقاس بالهرتز. وفي مشروعك، كل ملفات مجموعة MDVR-KCL مسجّلة بمعدّل **44100 هرتز**، أي 44100 قياس في كل ثانية من الصوت. المصدر: *docs/03_dataset.md* القسم *3.3*

مثال رقمي صغير. لنحسب حجم تسجيل واحد نموذجي في مشروعك:

- طول التسجيل تقريباً: دقيقتان، أي 120 ثانية.
- معدّل العيّنات: 44100 عيّنة في الثانية.
- الناتج: 120 × 44100 = **5,292,000 عيّنة**.

أي أن التسجيل الواحد في مشروعك ليس «صوتاً» بالنسبة للحاسوب، بل هو ببساطة **قائمة من حوالي 5.3 مليون رقم**. المصدر: *docs/04_pipeline.md* القسم *4.1*

وهذا الرقم يشرح لك أمرين عمليّين: لماذا استغرق استخراج الخصائص من كل الملفات حوالي 32 دقيقة، ولماذا يحتاج تحليل ملف واحد في التطبيق حوالي دقيقة كاملة.

خامساً: لماذا 44100 هرتز تحديداً؟

هذا سؤال ممتاز وقد يُسأل في المناقشة. والجواب يعتمد على قانون فيزيائي اسمه **نظرية نايكوِست** (the Nyquist theorem)، وتقول باختصار: **لكي تمثّل تردداً ما تمثيلاً أميناً، يجب أن يكون معدّل العيّنات ضعف ذلك التردد على الأقل.**

ولنفهم لماذا: تخيّل موجة تصعد وتنزل. إن أخذتَ عيّنة واحدة فقط في كل دورة، فقد تلتقط دائماً قمّة الموجة وتظنّها خطاً مستقيماً. تحتاج على الأقل قياسين في كل دورة، واحداً للقمّة وواحداً للقاع، لتعرف أنها تتذبذب أصلاً.

والتطبيق على مشروعك: معدّل العيّنات 44100 هرتز، إذن أعلى تردد يمكن تمثيله بأمانة هو 44100 ÷ 2 = **22050 هرتز**. وأذن الإنسان تسمع تقريباً حتى 20000 هرتز. فمعدّل 44100 هرتز يغطّي كامل مدى السمع البشري مع هامش بسيط. المصدر: *docs/04_pipeline.md* القسم *4.1*

وسنرى في الجزء 15 أن هذه النقطة بالذات كانت فخّاً خطيراً كاد يُفسد التحقّق الخارجي، لأن ملفاً بمعدّل 16000 هرتز **لا يمكن فيزيائياً** أن يحتوي أي صوت فوق 8000 هرتز.

سادساً: السعة

لو كان معدّل العيّنات يجيب على سؤال «متى نقيس؟»، فإن السعة (amplitude) تجيب على سؤال «**ماذا** قِسنا؟». السعة هي قيمة العيّنة الواحدة، وهي تمثّل شدّة الضغط في تلك اللحظة، أي ما تُدركه أذنك على أنه عُلُوّ الصوت في تلك اللحظة. وعادةً تُخزَّن هذه القيم كأرقام عشرية بين 1.0- و1.0+.

فمقطع من الصوت في الحاسوب يبدو هكذا فعلياً:

```
0.0021, 0.0834, 0.1520, 0.0912, -0.0334, -0.1201, ...
```

وكل خصائص مشروعك الـ74 مستخرَجة من قائمة أرقام كهذه، لا من شيء آخر.

مفهوم مرتبط: القَصّ. إذا كان الصوت أعلى من المدى الذي يستطيع الجهاز تسجيله، فإن القمم تُقصّ وتصير مسطّحة عند القيمة القصوى. وهذا يُسمّى القَصّ (clipping)، وهو تشويه حقيقي للإشارة يُفسد قياسات جودة الصوت لأنه يُدخل عدم انتظام مصطنَعاً لم تُنتجه الحنجرة. ولهذا فإن تطبيقك يفحص نسبة العيّنات المقصوصة ويُظهر تحذيراً إذا تجاوزت 0.1٪. المصدر: *docs/07_prototype.md* القسم *7.4*

سابعاً: القنوات

القناة (channel) هي تسجيل مستقلّ واحد. فأحادي (mono) يعني قناة واحدة، أي قائمة أرقام واحدة، وثنائي (stereo) يعني قناتين. وكل ملفات MDVR-KCL أحادية القناة أصلاً. ومع ذلك فإن أول خطوة في أنبوب المعالجة تُحوّل أي ملف إلى أحادي، لأن التطبيق قد يستقبل من المستخدم ملفاً ثنائياً. المصدر: *docs/03_dataset.md* القسم *3.3* و*docs/04_pipeline.md* القسم *4.2*

والسبب العلمي في اشتراط الأحادي: خصائصنا تصف **مصدراً صوتياً واحداً** هو الحنجرة. ولو حلّلنا قناتين مختلفتين قليلاً لحصلنا على قيمتين مختلفتين لنفس الشيء، وهذا لا معنى له.

ثامناً: ملف WAV ولماذا هو المهم هنا

صيغة WAV صيغة صوتية **غير مضغوطة** (uncompressed): تخزّن العيّنات كما هي، رقماً رقماً، دون أي فقدان في الجودة.

وهنا نقطة علمية مهمة جداً: صيغ مثل MP3 تُصغّر حجم الملف بأسلوب ذكي، إذ تحذف أجزاءً من الصوت **تظنّ أن الأذن البشرية لن تلاحظها**. وهذا مقبول تماماً للموسيقى، لكنه كارثة لمشروعك. لماذا؟ لأن ما تحذفه تلك الخوارزميات هو غالباً التفاصيل الخافتة وغير المنتظمة، وهي بالضبط ما نقيسه نحن. فعدم الانتظام الدقيق بين دورة وأخرى قد يكون غير مسموع للأذن، لكنه **هو الإشارة الطبية نفسها**.

لذلك كل ملفات مجموعتَي البيانات في مشروعك بصيغة WAV. ولهذا السبب نفسه رفض مشروعك لاحقاً إزالة الضجيج، وهي قصّة الجزء 4.

تاسعاً: الصورة الكاملة في سطر واحد

اهتزاز الحبال الصوتية يُنتج موجة ضغط في الهواء، ثم يحوّلها الميكروفون إلى جهد كهربائي متغيّر، ثم يقيس المحوّل ADC هذا الجهد 44100 مرة في الثانية، ثم تُخزَّن هذه القياسات كقائمة أرقام في ملف WAV، وهذه القائمة هي كل ما يراه برنامجك.

خلاصة الجزء

- العيّنة قياس واحد لشدّة الصوت في لحظة واحدة، ومعدّل العيّنات عددها في الثانية، وهو 44100 في مشروعك، أي أن تسجيلاً من دقيقتين قائمة من نحو 5.3 مليون رقم.

- اختيار 44100 هرتز يعود لنظرية نايكوِست: معدّل العيّنات يجب أن يكون ضعف أعلى تردد نريد تمثيله، و22050 هرتز تغطّي كامل السمع البشري.
- صيغة WAV غير المضغوطة ضرورية لأن الضغط يحذف التفاصيل الدقيقة غير المنتظمة، وهي بالتحديد الإشارة التي يقيسها المشروع.