

ELE3312 Projet final

Automne 2019

Version provisoire v1.0

Introduction

Le projet de cette année consiste à concevoir un système de détection et d'identification vocale. À l'aide d'un capteur à ultrasons et de deux microphones, vous devrez déterminer la position d'un de vos camarades. La détection se fera en deux temps. Tout d'abord, il faudra diriger (manuellement) un capteur à ultrasons en direction de la personne cible afin de **déterminer la distance** entre elle et le système. Par la suite, la personne visée devra prononcer un mot et vous devrez **mesurer la différence** de phase du son capté à l'aide de deux microphones séparés d'une distance connue. À l'aide de ces informations, il faudra afficher, en temps réel, **une flèche** qui pointe dans la direction du son émis. La dernière étape consistera à **identifier le mot** prononcé à l'aide d'un système de reconnaissance vocale.

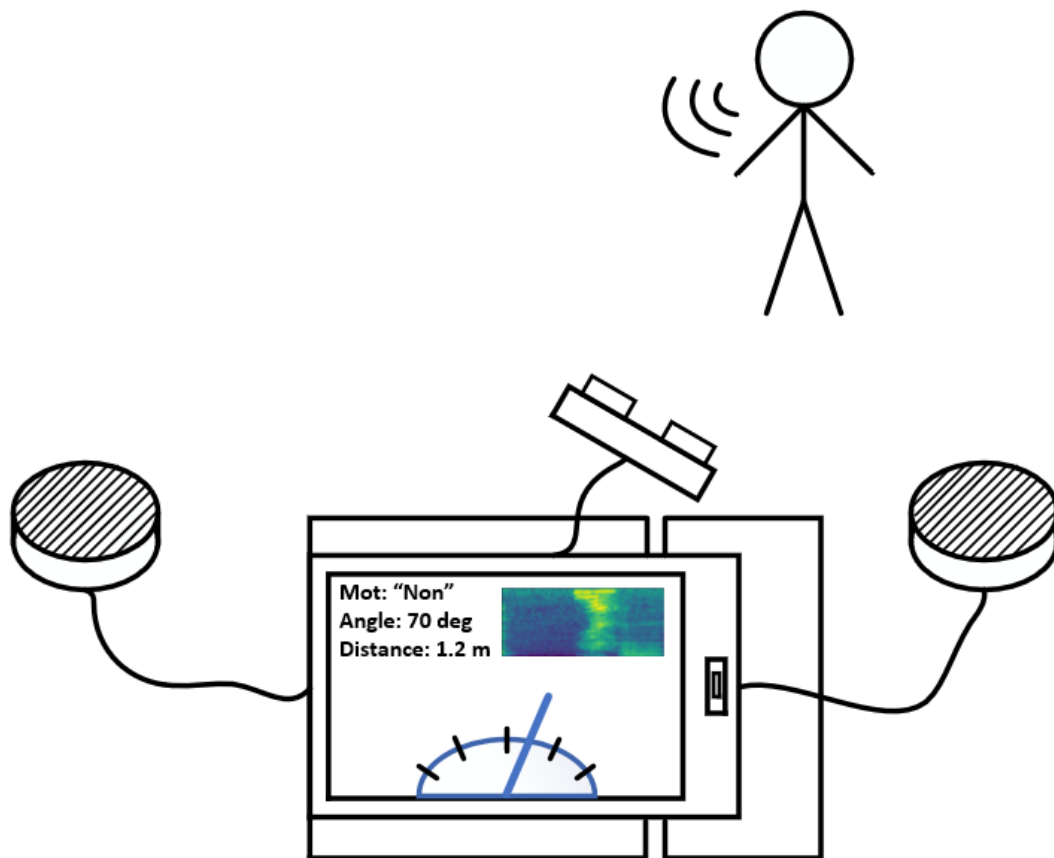


Figure 1 : Vue d'ensemble du projet. Les données ainsi que le spectrogramme de la voix doivent être affichés en temps réel.

Semaine 1 : Détection de la distance par ultrasons

Pour cette première partie (à présenter au début de la deuxième séance), vous devez **mesurer la distance** entre le capteur ultrasonore et la personne qui prononcera le mot/son clé. Le fonctionnement du capteur ainsi que les détails techniques sont présentés ci-dessous.

Fonctionnement du capteur

Le capteur à ultrasons possède 4 broches (**VCC, TRIG, ECHO et GND**). Pour son alimentation, les broches **VCC et GND** doivent être branchées sur du **5V et GND** respectivement. Vous pouvez vous servir des broches d'alimentation 5V et GND de votre microcontrôleur pour générer l'alimentation nécessaire. La broche **TRIG** permet de déclencher une mesure lorsqu'elle reçoit en son entrée une tension de **3.3V pendant 10 µs**. La broche ECHO envoie une impulsion de durée **proportionnelle à la distance** mesurée par le capteur. Le branchement des broches TRIG et ECHO doit se faire à l'aide des GPIOs de votre microcontrôleur configuré en entrée(s) et en sortie(s).

Attention, le signal ECHO est un signal 5V que vous devez ramener à un signal 3.3V au moyen d'un diviseur résistif sinon vous risquez d'endommager votre carte à microcontrôleur.

Afin de réaliser une conversion, les étapes suivantes sont nécessaires :

1. Émettre un signal **HIGH (3.3V)** sur la broche **TRIG pendant 10 µs**.
2. Attendre la détection du rebondissement des ultrasons émis par le capteur sur l'objet visé.
3. Détection d'un signal **HIGH** après le **court délai** de l'étape 2 sur la broche ECHO pour une durée égale au temps de vol des ultrasons. Ce temps doit être mesuré par le microcontrôleur.
4. Calculer la distance entre le capteur et l'objet selon la formule suivante :

$$\text{Distance en cm} = \text{temps de propagation en } \mu\text{s} / 58$$

Voici un diagramme temporel du dispositif :

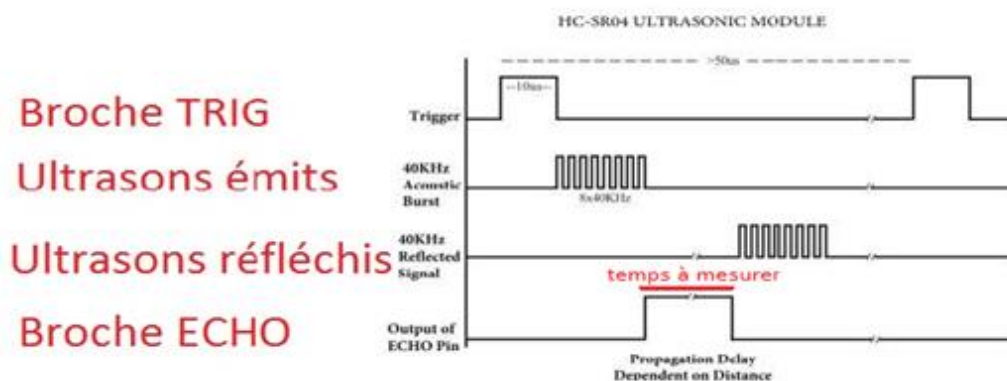


Figure 2 : Chronogramme des signaux du capteur ultrasonore

Détails techniques

Pour le début de la deuxième séance, vous devez être capable de mesurer la distance en temps réel et de l'afficher sur l'écran LCD avec un taux de rafraîchissement de **10 fois par seconde**. Le capteur devrait être en mesure de fournir une mesure sur une distance allant de **(3 cm à 2 m)**. Nous vous recommandons d'utiliser les broches **PB6 et PC8** pour les signaux **TRIG et ECHO**. Idéalement, vous utiliserez un **PWM pour le signal TRIG**.

Pour plus de détails Consulter :

<https://cdn.sparkfun.com/datasheets/Sensors/Proximity/HCSR04.pdf>

Semaine 2 : Mesure de la direction

Il est possible de déterminer la direction (l'angle de provenance de l'onde sonore) en mesurant le délai de propagation du son entre les deux microphones. Pour ce faire, nous vous conseillons d'utiliser une fonction de corrélation. Cette opération est généralement utilisée pour évaluer la ressemblance entre deux signaux. En fournissant deux ondes sonores, cette opération retourne un troisième signal dont le maximum représente le point de ressemblance maximal des deux signaux. En connaissant le taux d'échantillonnage ainsi que l'index de la valeur maximale de la fonction de corrélation, il est possible de déterminer le temps supplémentaire nécessaire pour atteindre le microphone le plus éloigné. Considérant que la vitesse du son dans l'air est d'environ 340 m/s, ce temps peut être traduit en distance.

La Figure 3 illustre un exemple d'une fonction de corrélation. Le signal en rouge se déplace graduellement vers la droite et le résultat (en noir) correspond au produit des aires partagées. La corrélation est une opération quasi-identique à la convolution. La seule différence étant que le signal d'entrée n'est pas inversé. L'aire partagée entre le signal bleu et rouge est maximale entre les étapes B et C (pic du triangle). Si le signal rouge est décalé par rapport au signal bleu, le pic maximal sera aussi décalé. C'est ce décalage qu'on cherche à mesurer.

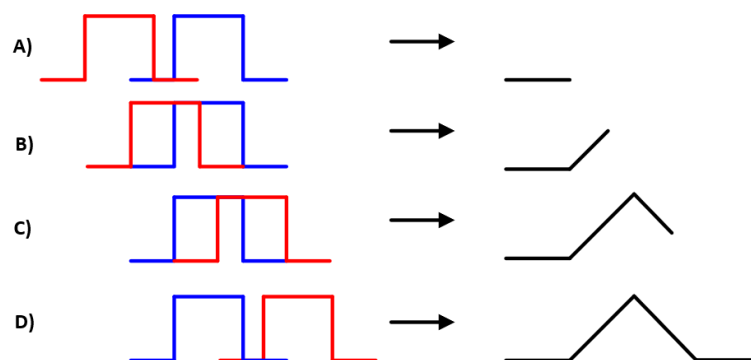


Figure 3 : Fonction de corrélation.

À l'aide de la distance obtenue par le capteur ultrasonore et de la distance calculée grâce au déphasage entre les signaux captés par les microphones, vous devez afficher à l'écran un curseur qui pointe dans la direction de la personne qui parle.

Une hyperbole peut être définie comme une courbe dont la différence de la distance entre un point sur la courbe et deux points fixes est constante. Dans notre situation, ces points fixes sont représentés par les microphones. La Figure 4 illustre ce concept.

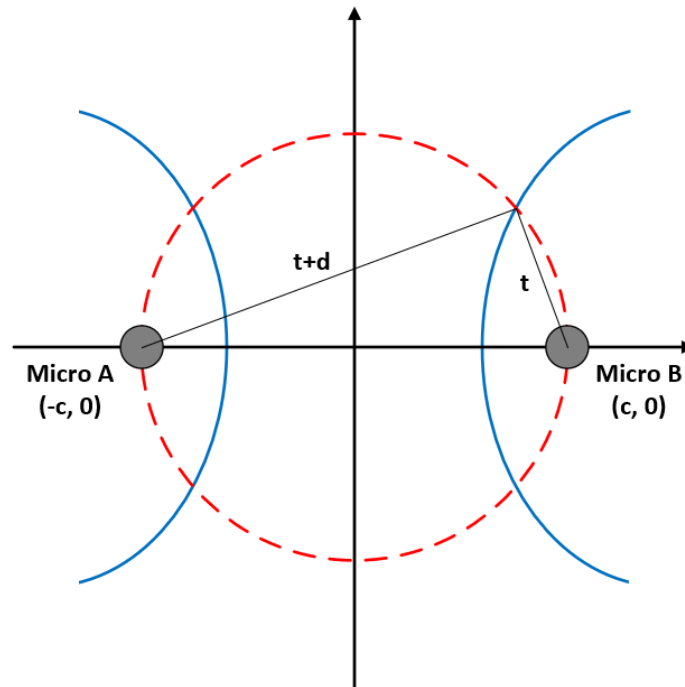


Figure 4 : Calcul de la position à l'aide d'une hyperbole (bleue) et d'un cercle (rouge). Les microphones sont séparés d'une distance $2c$ et le capteur à ultrasons se situe à l'origine.

Dans ce schéma, la valeur t correspond au temps requis pour atteindre le microphone le plus proche de l'interlocuteur. Le deuxième microphone a besoin d'un temps supplémentaire d . Puisqu'on ne connaît que la valeur d , on peut se retrouver à n'importe quel endroit sur la courbe en bleue. Afin de déterminer la position exacte de la personne qui parle, il faut aussi utiliser la valeur retournée par le capteur à ultrasons. La valeur retournée par le capteur correspond au rayon du cercle en rouge. Afin de vous aider, utilisez les équations suivantes :

Équations de l'hyperbole :

$$\frac{x^2}{a^2} - \frac{y^2}{c^2 - a^2} = 1$$

$$d = 2a$$

Où d correspond au temps supplémentaire requis pour atteindre le micro le plus éloigné et c la moitié de la distance entre les deux microphones.

Équation du cercle :

$$x^2 + y^2 = r^2$$

Où r correspond à la distance retournée par le capteur à ultrasons.

Fonctionnement des microphones

Les microphones possèdent 3 broches : OUT, GND et VCC. Les broches GND et VCC doivent être branchées sur GND et 3.3 V respectivement. La broche OUT doit être branchée directement à un ADC. Cette broche émet un signal audio centré autour de 1.65 V. Un petit potentiomètre situé en arrière du microphone permet d'ajuster le gain en tension. Afin de bien mesurer un déphasage entre les signaux captés par les microphones, on vous recommande de les placer à au moins 30 cm l'un de l'autre.



Figure 5 : Microphone fourni

Détails techniques

Pour la détection de l'onde sonore, vous devez utiliser une fréquence d'échantillonnage de 16 kHz. Vous devez utiliser des tableaux de 160 valeurs (10 ms d'audio) afin de comparer les deux signaux (ces tableaux seront aussi utilisés dans la partie 3). Il est aussi recommandé d'instancier un double buffer et de les remplir avec le DMA (comme dans le laboratoire 8). On vous recommande d'utiliser les broches PC2 et PC3 pour lire la tension des microphones.

Lien vers la librairie CMSIS-DSP (fonction de corrélation) :
http://www.keil.com/pack/doc/CMSIS/DSP/html/group_Corr.html

Semaine 3 : Détection de la voix

La troisième partie consiste à implémenter un système de reconnaissance vocale. En gros, le système doit être capable d'identifier le mot prononcé parmi une banque de mots prédéterminés. Lorsqu'un des mots de la banque est prononcé, le système doit afficher le mot, l'angle, la distance, le spectrogramme de la voix et doit pointer dans la direction de l'interlocuteur (voir Figure 1). Ces informations restent affichées pour trois secondes. Par la suite, le système se remet en mode détection et pointe dans la direction du son courant tant qu'un autre mot n'est pas prononcé. Le scénario suivant illustre l'ordre des opérations :

1. Un des coéquipiers (A) se place devant le système à un endroit de son choix.
2. Un deuxième coéquipier (B) pointe le capteur ultrasonore dans la direction du coéquipier A. L'écran LCD affiche le spectrogramme courant, la distance et le curseur pointe en direction du son émis.
3. Le coéquipier A prononce un des mots de la banque.

4. Le système détecte sa position et le mot prononcé. Le mot, l'angle, la distance, le spectrogramme et le curseur figent pour 3 secondes. Le curseur doit pointer dans la direction du coéquipier A.
5. Le système recommence et l'écran LCD affiche à nouveau le spectrogramme, la distance et le curseur. Ces valeurs sont mises à jour continuellement.

Les systèmes de reconnaissance vocale modernes utilisent généralement des algorithmes complexes découlant de l'apprentissage machine. Nous ne vous demandons pas de reproduire ces algorithmes. Toutefois, il est possible, avec vos connaissances actuelles, de produire un système relativement précis. Pour ce faire, nous vous fournirons des fonctions pour calculer le Mel-spectrogramme, soit une représentation utilisée pour ce genre de tâche.

Notes concernant le spectrogramme :

Afin de calculer le spectrogramme, vous devez utiliser une fréquence d'échantillonnage de 16 kHz. Vous avez, au préalable, comparé des tableaux de 160 points (représentant 10 ms d'audio). En concaténant trois de ces tableaux (30 ms), vous pouvez calculer les coefficients représentant le Mel-spectrogramme, typiquement utilisé pour des systèmes de reconnaissance vocale. Ces coefficients, au nombre de 40, sont calculés à chaque 10 ms. La Figure 6 illustre la procédure générale.

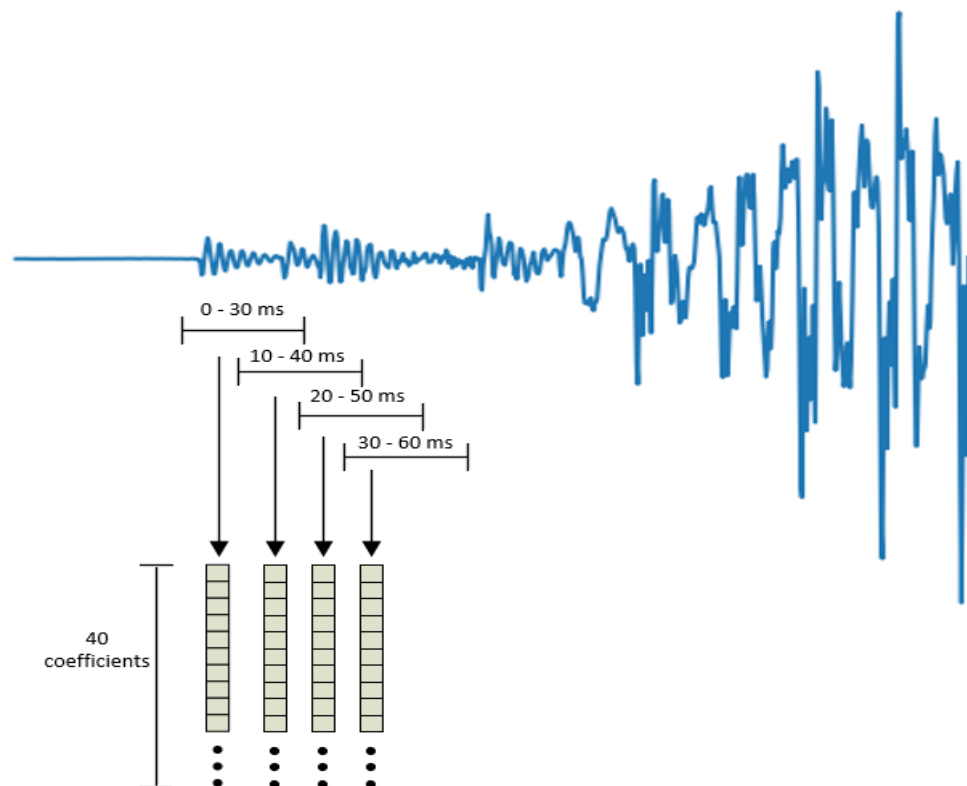


Figure 6 : Calcul des coefficients du spectrogramme. Chaque fenêtre de 30 ms retourne 40 coefficients. Des nouveaux coefficients sont calculés à chaque 10 ms.

Pour le projet, nous vous fournissons la fonction, dont le prototype est :

```
void compute_log_mel_coefficients(float* pSrc, float* pDst);
```

Cette fonction prend en entrée un tableau de 160 valeurs représentant les derniers 10 ms captés. Elle retourne un pointeur vers un tableau de 40 valeurs représentant les coefficients du spectrogramme calculés pour les 30 ms précédentes (les 20 ms de l'appel précédent sont sauvegardées dans un tableau).

Afin de vérifier l'implémentation de votre système, nous vous demandons d'afficher le spectrogramme sur l'écran LCD pour une séquence audio de 1 seconde. Le spectrogramme devrait ressembler à celui retrouvé à la Figure 7. Pour obtenir cette image, il suffit de concaténer les 40 coefficients que vous obtenez pour chaque tranche de 30 ms. Puisque vous prenez une nouvelle tranche à chaque 10 ms, vous obtiendrez environ 100 fois 40 coefficients. Il faut dessiner les coefficients au fur et à mesure qu'ils sont calculés. Lorsque vous dessinerez le spectrogramme à l'écran il faudra utiliser au moins 4 couleurs différentes afin de représenter les différentes intensités de pixels.

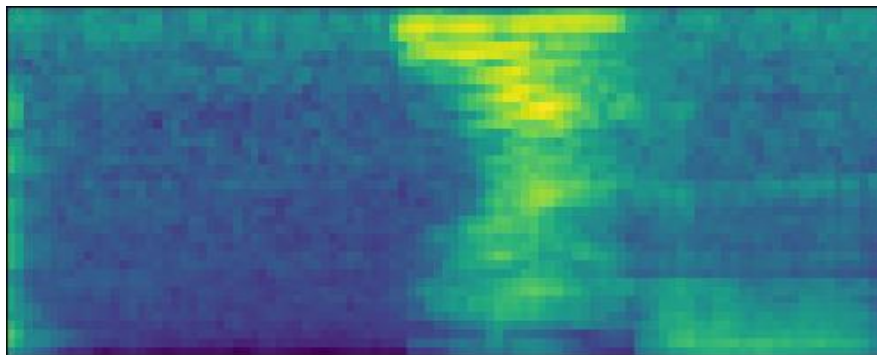


Figure 7 : Mel-spectrogramme du mot "Yes". Les dimensions de l'image sont de 40 x 100 pixels.

Après avoir calculé le spectrogramme, vous devrez trouver une manière de détecter le son qu'il représente. Pour cette étape, vous êtes laissés par vous-mêmes. À vous de trouver une façon ingénieuse de détecter les mots. Bonne chance !

Détails techniques :

Lors de l'évaluation au laboratoire, nous vous demandons de reconnaître 4 mots de votre choix prononcés par la personne de votre choix également, par exemple : "gauche", "droite", "haut", "bas". Le système ne doit pas générer de faux positifs (mot détecté si rien n'est prononcé).

Pondération (16 points au total)

3 points pour la mesure ultrasonore (Pour le début de la séance 2) :

1. Précision de la mesure
 - o 2pt : on détecte la distance correctement (1pt si résolution < 5cm)
2. La mise à jour en continu
 - o 1pt : on affiche en continu la distance mesurée (rafraîchie aux 100 ms)

4 points pour la détection de l'angle (Pour le début de la séance 3) :

1. Mesure de la tension générée par les micros
 - o 1pt : Les deux ADCs sont configurés correctement
2. Utilisation d'un double tampon et du DMA
 - o 1pt : Les tableaux sont automatiquement remplis/des tampons sont utilisés
3. Mesure du délai et calcul de l'angle (direction)
 - o 2pt : L'angle est affiché correctement à l'écran
 - o 1.5pt : Le délai entre les micros et la distance (ultrasons) sont mesurés, mais la direction n'est pas calculée.
 - o 1pt : Le délai entre les micros est mesuré, mais n'est pas précis.

4 points pour l'application complète (Pour la séance 4) :

1. Fonctionnelle
 - o 1pt : Le système est capable de pointer dans la direction du son émis
 - o 0.5pt : Le système est parfois capable de pointer dans la bonne direction
2. Calcul du spectrogramme
 - o 1pt : Le spectrogramme s'affiche correctement.
 - o 0.5pt : Le spectrogramme n'est pas en temps réel ou est incomplet.
3. Capacité à détecter des mots
 - o 2pt : Le système est capable d'identifier 4 mots différents.
 - o 1.5pt : Le système est capable d'identifier 3 mots différents.
 - o 1pt : Le système est capable d'identifier 2 mots différents.
 - o 0.5pt : Le système peut détecter au moins 1 mot.

2 points pour la présentation vidéo technique (2 minutes maximum) :

- o 1pt : La vidéo illustre tous les points évalués ci-dessus
- o 1pt : La vidéo démontre le bon fonctionnement du projet.

Cette vidéo doit être remise sur moodle pour des fins d'évaluation. Dans cette vidéo, vous devez présenter le fonctionnement complet de votre projet dans les meilleures conditions possibles. Si votre projet n'a pas fonctionné le jour de l'évaluation au laboratoire, ceci est votre chance d'obtenir quelques points.

3 points pour la démonstration

- o 3pts : Selon une version simplifiée du suffrage par jugement majoritaire (voir ci-dessous)

La compétition

Le jour de la compétition, vous devrez présenter une vidéo de type "marketing" qui durera environ 1 minute afin de vendre les mérites du projet. Expliquez comment vous avez réussi à détecter des mots.

Comme il y aura environ 30 équipes de 3-4 étudiants, chaque équipe aura droit à maximum 1 minute de vidéo et maximum 3 minutes de démo en avant avec 1 minute de battement entre deux équipes, afin d'avoir 10 équipes par heure.

Le but de la compétition est de pouvoir détecter le plus de mots possibles. Chaque équipe aura le droit à deux erreurs. Dans les deux cas, vous perdez une chance. L'équipe qui réussira à détecter le plus de mots différents avant de passer à travers les deux chances gagnera la compétition et obtiendra 1 point boni.

Vous serez évalués par les autres équipes au moyen d'une version simplifiée du suffrage par jugement majoritaire. En gros, vous devrez noter chaque équipe avec une mention parmi les suivantes :

- ✓ Excellent (3pt)
- ✓ Très bien (2pt)
- ✓ Bien (1pt)
- ✓ Passable (0pt)
- ✓ Insuffisant
- ✓ Très insuffisant

Votre note sera celle qui rallie minimalement 50% des suffrages. Par exemple, si vous avez recueilli les votes suivants :

5%	Excellent
25%	Très bien
30%	Bien
20%	Passable
15%	Insuffisant
5%	Très insuffisant

Au moins 50% des étudiants pensent que vous méritez "Bien" ou plus ... et en fait, c'est même 60%

Soyez bons joueurs ! Je me réserve le droit de supprimer des voteurs ou des votes trop éloignés de la tendance ou encore d'ajuster les notes pour parer à toute situation imprévue.

Que le meilleur gagne !!!