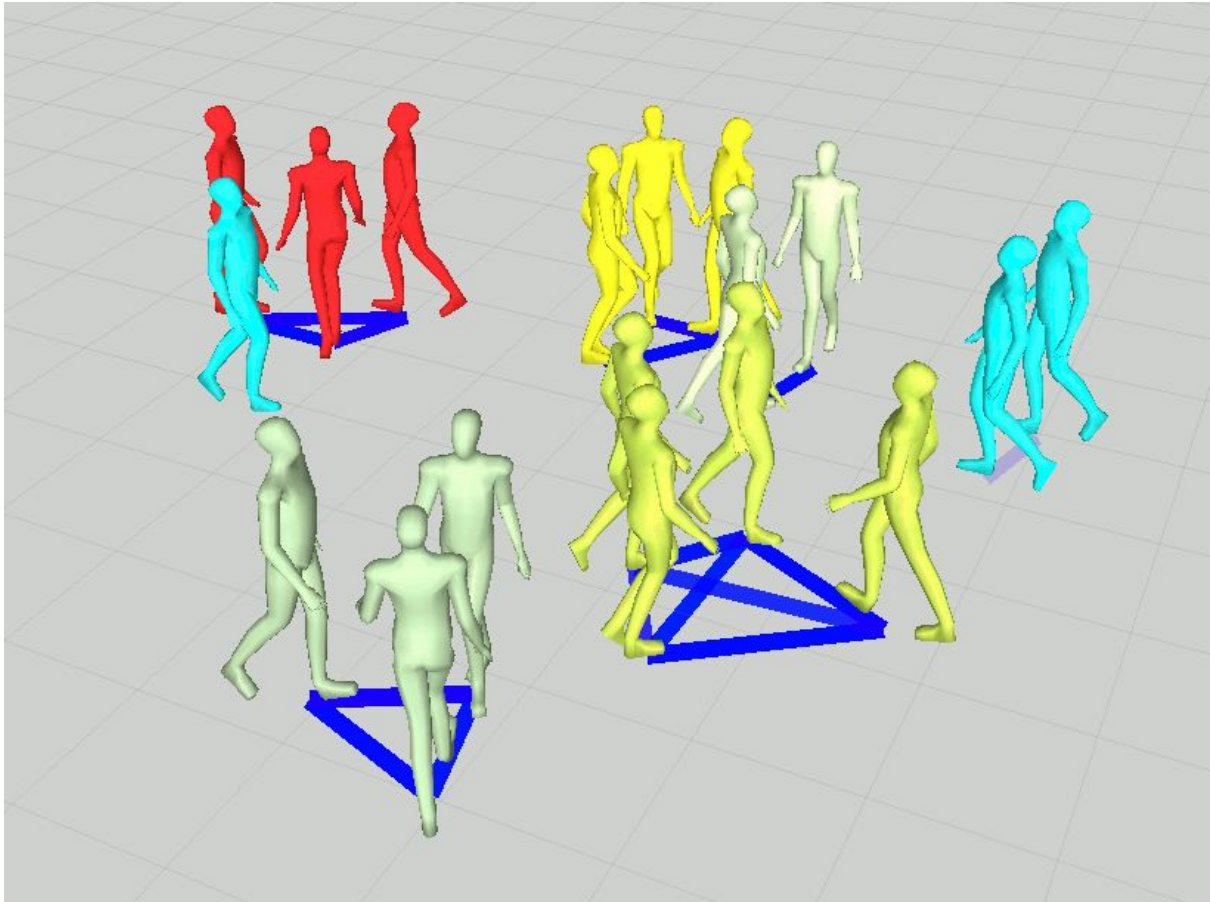


# Modeling Interpersonal Interactions in Multi-Party



Julien Defiolles

M2 ANDROIDE  
Sciences Sorbonne Université



Encadrants: Mohamed Chetouani, Catherine Pelachaud  
Responsable de stage: Nicolas Bredeche

# Sommaire

<b>Sommaire</b>	<b>1</b>
<b>I) Introduction</b>	<b>2</b>
Les motivations	2
Les objectifs du stage	2
Le cheminement du stage	3
<b>II) État de l’art</b>	<b>3</b>
<b>III) Le corpus ‘council of coaches’</b>	<b>5</b>
<b>IV) L’extraction, analyse et calculs des différents paramètres</b>	<b>6</b>
L’extraction automatique des paramètres de prosodie	6
Le calcul des paramètres pour la synchronie	6
Le temps de pause	7
Le temps de parole	7
Le silence	7
Les backchannels	8
L’Analyse	9
<b>V) Le calcul de la synchronie</b>	<b>10</b>
Introduction	10
Les différentes parties du calcul de la synchronie interpersonnelle	11
Le taux de latence	11
Le temps de pause	12
Le temps de parole	13
Le silence	14
Le calcul final de la synchronie interpersonnelle	15
Les tests	15
La synchronie sur différentes sessions	16
La synchronie en cas d’ajout ou suppression de personne	17
Le test du calcul de la synchronie	18
<b>VI) Conclusion et suite</b>	<b>21</b>
<b>Bibliographie</b>	<b>21</b>

# I) Introduction

## 1) Les motivations

Il y a un intérêt grandissant pour l'analyse, la modélisation et la synthèse automatique de comportements non-verbaux au niveau d'un groupe d'agents virtuels. Chaque personne a un ensemble de comportements dicté entre autre par des normes sociales ou son éducation, qui influence sa façon d'interagir avec d'autres personnes, en modifiant ses gestes et sa parole. Plusieurs modèles ont été proposés pour gérer les tours de parole et les interruptions dans une conversation entre plusieurs humains. Dans un groupe, les humains se positionnent et prennent la parole en fonction de leurs intentions et de leurs attitudes sociales. L'interaction interpersonnelle peut être calculée par l'analyse automatique des signaux vocaux et visuels échangés entre partenaires. Le concept de synchronie interpersonnelle a été utilisé pour modéliser les interactions dyadiques et de groupe. La synchronie interpersonnelle étant une valeur qui représente la coordination temporelle durant les interactions sociales, cela implique des échanges fluides et dynamiques entre les différentes parties qui communiquent (Delaherche et al. 2012). Un certain nombre de quantificateurs ont été proposés pour caractériser différents aspects de la synchronisation entre partenaires au sein d'une interaction, ainsi que l'interaction elle-même. Les mesures de synchronie peuvent servir à quantifier l'interaction interpersonnelle.

La plupart des travaux sur l'interaction humain-agent portent sur l'interaction dyadique. Cependant, rares sont ceux qui abordent l'interaction de groupe impliquant plusieurs agents ou des humains (dans ce rapport on parle de groupes allant de 3 à 15 personnes ou agents). Parmi les travaux existants, on peut noter les modèles de gestion de la prise de parole (Ravenet et al. 2015), les modèles du regard (Oertel et Salvi, 2013), la gestion du dialogue (Traum et al., 2012) et les compétences sociales (Prada et Paiva, 2009). Très peu de travaux s'intéressent à la modélisation des comportements des agents sociaux en interaction avec d'autres agents ou humains.

## 2) Les objectifs du stage

L'objectif de ce stage est de développer un modèle d'interaction de groupe pour des agents sociaux. Cet objectif est découpé en trois parties plus distinctes:

1. Analyser des modèles d'interaction interpersonnelle émergent dans l'interaction multi-partie humaine.
2. Proposer de nouvelles mesures pour calculer la synchronisation interpersonnelle dans un groupe d'agents humains.
3. Mettre en œuvre le modèle appris de synchronisation interpersonnelle dans l'interaction multi-partie des agents virtuels.

Au cours des 6 mois de stage, les étapes 1 et 2 ont pu être réalisées. Tout d'abord, en découvrant les différents travaux déjà effectués sur les interactions interpersonnelles dans un groupe d'agents, autant dans les travaux de type informatique concernant plus spécifiquement l'analyse et l'évaluation de ces interactions que dans les travaux de type

psychologique portant sur l'explication de ces comportements. Ensuite, avec cette compréhension du fonctionnement d'un groupe, de son organisation et l'analyse des différentes méthodes déjà existantes pour mesurer la synchronie interpersonnelle, on peut proposer une toute nouvelle mesure de la synchronie de groupe à partir de différents paramètres vocaux. La troisième partie n'a pas pu être réalisée à cause d'un manque de temps.

### 3) Le cheminement du stage

Afin d'obtenir un calcul de la synchronie interpersonnelle de groupe, il a fallu au préalable apprendre les différentes définitions des termes du sujet et connaître les différents travaux déjà effectués dessus. Ainsi la lecture de différents articles et livres portant sur la psychologie de groupe, sur les différentes méthodes informatique ou psychologique pour calculer la synchronie, fût le début du travail.

Pour travailler sur la synchronie interpersonnelle et obtenir des résultats pour effectuer des tests dessus par la suite, il fallait un corpus de vidéo. Le choix fût d'en prendre un appartenant déjà à l'équipe: le corpus 'council of coaches'.

Par la suite, il a fallu chercher les différents paramètres pour calculer la synchronie, on a pris des paramètres audio pour commencer, le choix a été fait sur le temps de pause, le temps de parole, le silence, le taux de latence et les backchannels. Un backchannel est une courte utterance produite par une personne de la conversation pendant qu'une autre personne parle pour montrer de l'intérêt. Afin de calculer ces différents paramètres, il fallait un outil pour extraire les paramètres de prosodies des vidéos du corpus, permettant de calculer les données évoquées plus haut. Pour cela on a utilisé le logiciel open source OpenSmile.

A partir des différents paramètres calculés, il a fallu mélanger les différents paramètres pour obtenir une valeur de la synchronie interpersonnelle.

Une fois le calcul mis au point, il ne restait qu'à tester les résultats sur différents points:

- La synchronie calculée effectue-t-elle les bonnes variations de valeurs aux bons moments ?
- La synchronie est-elle indépendante du nombre de personnes et de la vidéo où celle-ci est calculée ?
- Le calcul représente-t-il une bonne approximation de la synchronie ?

## II) État de l'art

Il y a peu de travaux portant sur les interactions de groupes composés d'humains et/ou d'agents virtuels comme nous le définissons, c'est-à-dire des groupes de 3 à 15 personnes ou agents. On peut en observer certains comme celui portant sur Furhat (Skantze, Johansson, et Beskow 2015), un robot discutant avec 2 personnes pour gagner un jeu. Il y a aussi des travaux portant sur l'analyse de dialogues en groupes, par exemple le papier *Incremental dialogue understanding and feedback for multiparty, multimodal conversation* (Traum et al. 2012) explique une architecture pour permettre à un agent de

générer des feedbacks au bon moment à partir d'une conversation de groupe en temps réel

On constate aussi qu'il y a déjà des travaux sur la synchronie interpersonnelle. Cette dernière étant une valeur représentant la coordination temporelle durant les interactions sociales, cela implique des échanges fluides et dynamiques entre les différentes parties qui communiquent (Delaherche et al. 2012). Mais la plupart de ces travaux sont sur des dyades, c'est-à-dire avec deux agents humains ou virtuels. Ainsi dans l'article *Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines* (Delaherche et al. 2012) on nous explique différentes manières de calculer la synchronie interpersonnelle à partir de travaux issus de la psychologie ou de l'informatique en dyade. Certaines des méthodes pour calculer la synchronie sont déjà implémentées dans le logiciel SyncPy (Varni et al. 2015), une librairie open source permettant de calculer entre autre la synchronie, mais pour deux personnes au maximum. Ces travaux ne sont pas applicable pour un groupe humain ou agents virtuels du fait de leurs modélisation. Ils prennent par exemple en compte le temps de réponse d'une personne à une autre, ce qui est simple avec un duo : on regarde lorsque l'autre personne répond. Mais c'est moins évident pour un groupe, il faut pouvoir savoir dans le groupe qui sera la personne répondant à l'orateur.

Pour aider au calcul de la synchronie en groupe, il a fallu récolter des informations sur la psychologie de groupe et son fonctionnement. Ainsi le livre *Nonverbal Communication* (Burgoon, Guerrero, et Floyd 2016) explique l'importance des comportements non verbaux au sein de la communication entre humains. Le livre *Group dynamics, the psychology of small group behavior* (Shaw 1971) montre les différents comportements d'un groupe de personnes en fonction de plusieurs facteurs comme son environnement physique, sa composition et sa structure.

Finalement, il y a aussi des recherches liées à des paramètres proches de la synchronie comme la cohésion. Carron (1982) définit la cohésion de groupe comme "un processus dynamique qui se caractérise par la tendance d'un groupe à se serrer les coudes et à demeurer unis dans la poursuite de ses objectifs". La cohésion est ainsi une mesure proche de la synchronie interpersonnelle car un groupe qui poursuit les mêmes objectifs va avoir tendance à être plus efficace et dynamique sur les différents échanges entre les personnes du groupe. Ainsi le papier *Estimating Cohesion in Small Groups Using Audio-Visual Nonverbal Behavior* (Hung et Gatica-Perez 2010) utilise des paramètres audio-visuels pour obtenir la cohésion de groupe, ces paramètres pouvant être réutilisés pour la synchronie interpersonnelle.

Il n'y a pas dans les recherches actuelles de méthode pour calculer la synchronie interpersonnelle dans un groupe humain. L'objectif est donc de créer une nouvelle méthode, à partir de méthodes calculant d'autres paramètres comme la cohésion et à partir des calculs de synchronie dyadique, afin d'obtenir la synchronie interpersonnelle en groupe.

### III) Le corpus 'council of coaches'

Pour extraire les différentes données et calculer la synchronie interpersonnelle, il fallait des vidéos support, pour cela on s'est tourné vers le corpus 'Council of coaches'.

Le projet 'Council of coaches' (logo figure 1) est un projet de recherche européen visant à créer un groupe de conseillers virtuels autonomes pouvant aider les gens à atteindre leurs objectifs de santé.



Figure 1: Logo du projet Council of coaches

Le corpus est composé de 7 séances d'environ 20 minutes, chacune consistant à conseiller une personne sur certains de ses problèmes de santé. On retrouve ainsi à chaque séance au départ Alan et Gauvin (étiquetés respectivement Blue et Yellow dans ce rapport) discutant en premier lieu du problème de la patiente (Red). Ils la font entrer dans la pièce et la conseillent sur son problème. Lorsqu'ils ont fini de discuter avec elle ils la font sortir, puis Yellow et Blue font un débriefing de ce qui s'est passé (exemple figure 2). Pour chaque séance du corpus, on possède la piste vidéo et piste audio de chacun des participants, ainsi que la vidéo et l'audio de la séance entière.

Ce corpus a été choisi car il appartient déjà à l'équipe du laboratoire et il a été annoté sur la cohésion par la méthode 'Thin Slice' (Ambady et Rosenthal 1992). Pour chaque fenêtre de 2 minutes de vidéo, 10 questions ont été posées, chacune ayant une réponse de 1 à 7, 1 étant 'pas du tout d'accord', 4 étant 'neutre' et 7 'tout à fait d'accord'.



Figure 2: Alan/Blue en bleu, Gauvin/Yellow en jaune et Red en rouge durant la session 1

## IV) L'extraction, analyse et calculs des différents paramètres

### 1) L'extraction automatique des paramètres de prosodie

Pour calculer la synchronie interpersonnelle à partir des paramètres audio, il faut commencer par extraire les paramètres de prosodie du corpus 'Council of coaches' afin de calculer par la suite des paramètres plus avancés et obtenir à partir de ces-derniers la synchronie. Ces données sont choisies afin de pouvoir déterminer la plupart des paramètres par la suite. Ainsi elles correspondent :

- Au paramètre IsTurn, qui est un booléen mis à 1 lorsque la personne parle et à 0 dans le cas contraire. Ce paramètre est utile pour calculer les paramètres comme le temps de pause, le temps de parole, le silence et le taux de latence.
- Aux paramètres d'énergie, d'intensité, de volume et hauteur sonores qui sont des flottants, correspondant aux différentes valeurs de puissance du son. Ils permettent quant à eux de trouver les backchannels.

Pour récupérer ces données, le plus simple était d'utiliser un logiciel déjà existant. Le choix a été de prendre le software openSMILE car c'est un logiciel open source qui a déjà été utilisé au sein de l'équipe du laboratoire. SMILE est l'acronyme pour Speech and Music



Interpretation by Large-space Extraction. L'outil d'extraction OpenSMILE permet de récupérer une large combinaison de paramètres audio en temps réel et de faire du traitement de langage. OpenSMILE prend en entrée un fichier de configuration décrivant les différentes données à sortir et les différents paramètres à prendre en compte pour l'extraction de celles-ci. Il sort, après exécution, les résultats voulus dans un fichier de type csv où chaque ligne représente l'ensemble des données voulues pour une frame de la vidéo.

### 2) Le calcul des paramètres pour la synchronie

L'ensemble des paramètres de prosodie calculés nous permet de nous pencher sur des paramètres plus avancés. L'objectif étant la synchronie interpersonnelle, il fallait trouver les bonnes données pour la calculer. On a donc suivi un papier calculant la cohésion (Hung et Gatica-Perez 2010) pour récupérer les différentes données calculables afin d'obtenir la cohésion. Cette valeur est très corrélée avec la synchronie interpersonnelle du fait de leurs ressemblances, on peut donc utiliser les paramètres de la cohésion pour obtenir une valeur de la synchronie interpersonnelle. On a ainsi extrait quatre différentes mesures pour les utiliser par la suite dans le calcul de la synchronie: le temps de pause, le silence, le temps de parole, et les backchannels. Tous les paramètres sont calculés sur une fenêtre de temps



choisie par l'utilisateur, ainsi celui-ci aura une valeur par paramètre sur la fenêtre de temps donnée.

### a) Le temps de pause

Le temps de pause est la somme de toutes les pauses dans une fenêtre durant le tour de parole d'une personne. Une pause est un intervalle de silence de 4 secondes maximum sur un tour de parole. Les 4s ont été arbitrairement choisis après plusieurs essais de différents temps. Ce temps permet d'éviter de prendre en compte les moments où une personne ne parle pas parce qu'elle attend qu'une personne lui réponde. Par exemple sur la figure 3, Bob a un temps de pause de 2s.

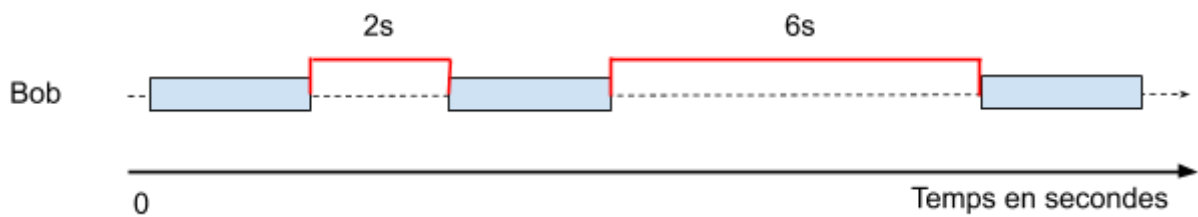


Figure 3: Temps de pause pour Bob.

Pour calculer le temps de pause il suffit de prendre le paramètre `isTurn`, si celui-ci est à zéro pendant moins de 4s, alors on ajoute la pause au temps de pause totale.

### b) Le temps de parole

Le temps de parole est la somme des temps des intervalles où l'un des membres du groupe parle sur une fenêtre donnée. Par exemple sur la figure 4, le temps de parole est de 10s.

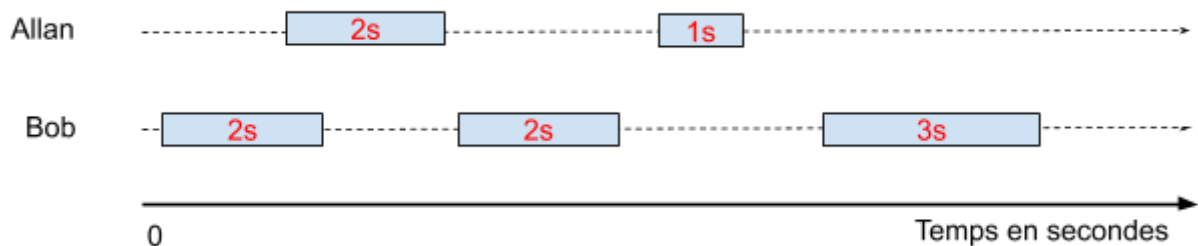


Figure 4: Temps de parole pour Bob et Alan.

Pour le calculer, on a pris le paramètre `isTurn` pour toutes les personnes du groupe et le temps de parole représente le nombre de fois où le paramètre `isTurn` est à 1, multiplié par le temps de la frame du `isTurn`.

### c) Le silence

Le silence est un intervalle dans une conversation où aucun membre du groupe ne prend la parole. Par exemple sur la figure 5, le temps de silence est de 3s.



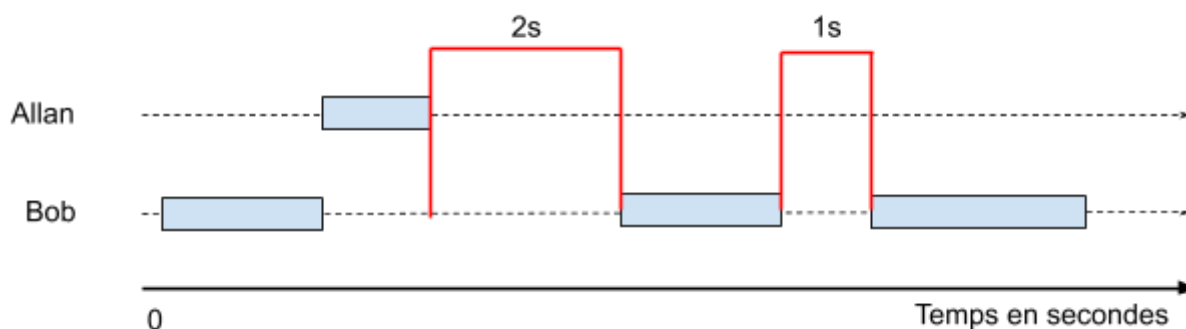


Figure 5: Le silence pour Bob et Alan.

Pour calculer le silence on a pris le paramètre isTurn comme pour les autres paramètres et à chaque fois que l'ensemble des toutes les personnes du groupe ont leurs paramètres isTurn à zéro, alors, on ajoute le temps de la frame au temps de silence.

#### d) Les backchannels

Un backchannel est une courte utterance produite par une personne de la conversation pendant qu'une autre personne parle pour montrer de l'intérêt, de l'attention et/ou une volonté de continuer à écouter. Par exemple les backchannels les plus courants dans la langue française sont oui, ouais, humhum, hum et ah. Les backchannels dans la synchronie interpersonnelle sont importants car ils montrent que les personnes sont impliquées dans l'échange, ce qui est important dans la fluidité de l'échange donc la synchronie. Pour trouver les différentes utterances qui sont des backchannels, on est partie d'un article (Ruede et al. 2017), qui utilise les paramètres de prosodie comme l'énergie, l'intensité et la hauteur du son dans un algorithme de deep learning. Cet algorithme prend en entrée l'énergie, l'intensité et la hauteur du son d'une utterance qui contient les mots correspondant aux backchannels, et donne en sortie un boolean indiquant si l'utterance est un backchannel ou pas. On est donc aussi partie d'un algorithme de deep learning qui prend en entrée les mêmes paramètres sur toutes les utterances plus la durée de l'utterance, car les backchannels ne durent pas longtemps et donne la même sortie. L'objectif était de partir de cet algorithme et de l'améliorer par la suite. Les résultats de cette recherche de backchannel sont présentés en table 1.

	Backchannel selon le classifieur	Non backchannel selon le classifieur	Totaux
Backchannel	12	21	33
Non backchannel	15	86	101

Table 1: Classification des backchannels avec un réseau neuronal appris sur 100 utterances (les non backchannels sont les utterances qui ne sont pas des backchannels)

Les résultats montrent que le réseau de neurones ne trouve que 38% des backchannels et que 56% des résultats sont faux, ce qui rend les résultats peu intéressants. Ces résultats sont dus à la très faible quantité de données. En effet, 100 utterances c'est peu, de plus la méthode sur lequel on se base utilise et entraîne son algorithme de deep

learning sur les utterances qui contiennent les mots correspondant aux backchannels. On ne pouvait pas faire ça à partir de notre corpus puisqu'on n'avait pas de moyen de récupérer rapidement ce que les acteurs disent. Nos résultats n'étant pas optimaux et par manque de temps, nous avons abandonné l'utilisation des backchannels dans le calcul de la synchronie.

### 3) L'Analyse

Pour tester les différents paramètres et voir s'ils correspondaient bien aux valeurs voulues, on a d'abord affiché le détail de qui parlait et quand. Sur la figure 6, on peut ainsi voir qu'entre 200s et 280s, Bob et Allan parlent puis c'est au tour de Linda et Bob.

Pour mieux analyser les différents paramètres, on a fait deux types de normalisation. La première consiste juste à normaliser les valeurs selon leur maximum et leur minimum. La seconde normalise en fonction des autres participants. On a aussi deux types d'affichages: un histogramme comme sur la figure 7 et un affichage pour nous montrer le pourcentage d'un paramètre d'une personne par rapport aux autres personnes présentes comme en figure 8.

Grâce à ces figures et aux vidéos, on a pu ainsi vérifier l'exactitude de nos paramètres.

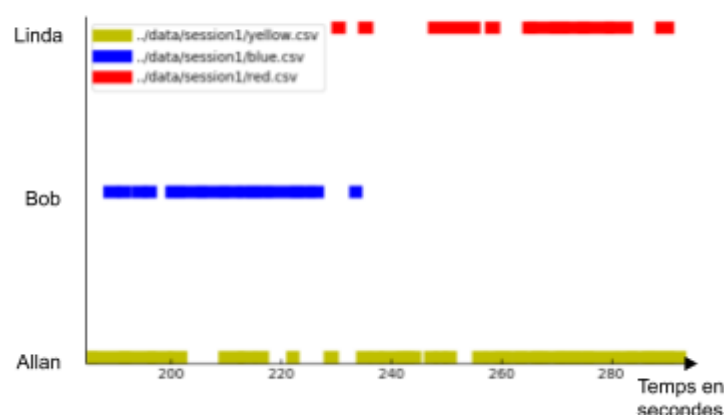


Figure 6: Les prises de parole entre Allan, Bob et Linda entre la 180 et la 300ème seconde.

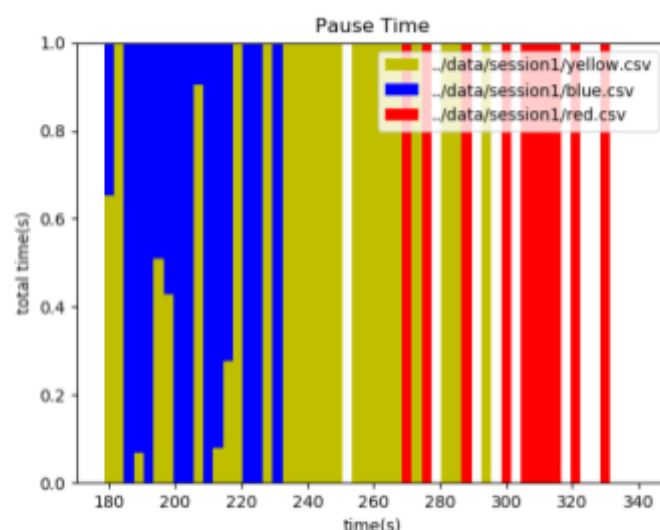


Figure 8: Temps de pause normalisé de la seconde façon, entre la 180 et la 340 ème seconde.

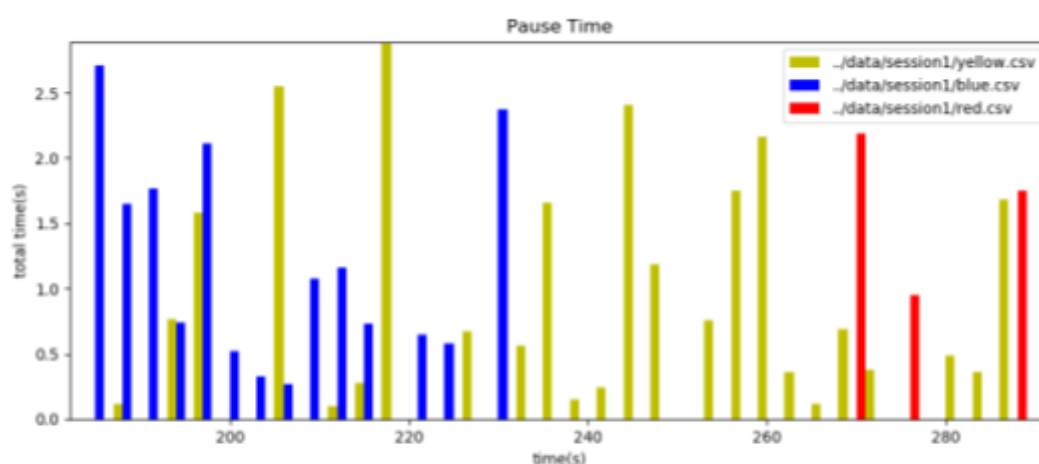


Figure 7: Histogramme du temps de pause, entre la 180 et la 300 ème seconde.

## V) Le calcul de la synchronie

### 1) Introduction

L'objectif de cette partie est de présenter une méthode pour calculer la synchronie interpersonnelle dans un groupe de personnes. Pour rappel, la synchronie interpersonnelle est la capacité pour plusieurs personnes d'être en phase, ce qui implique des échanges fluides et dynamiques.

Tous les résultats observés dans cette partie sont fait à partir du corpus 'Council of coaches' présenté plus haut.

Il existe déjà des calculs de synchronie interpersonnelle, mais seulement pour des échanges entre deux personnes. On peut retrouver ces méthodes dans le logiciel SyncPy. La méthode qui nous intéresse dans SyncPy est la méthode qui utilise juste le composant IsTurn. Elle consiste à calculer le temps de latence entre la prise de parole entre deux personnes, à calculer le nombre de fois où chaque personne a un temps de latence inférieur à un seuil, et de le diviser par le nombre total de réponses. On appelle cette mesure le taux de latence. Par exemple, sur la figure 9, le taux de latence avec un seuil de 1s pour Bob sur Allan est de 0.5 car il répond deux fois à Allan, avec une fois une latence de 0.5 et une fois une latence de 2s. Pour Allan sur Bob, le taux de latence est de 1. On est donc parti de cette méthode et on l'a étendue pour calculer la synchronie interpersonnelle. Le temps de latence n'étant pas suffisant, on a par la suite augmenté ce résultat avec le temps de pause, le temps de parole et le silence.

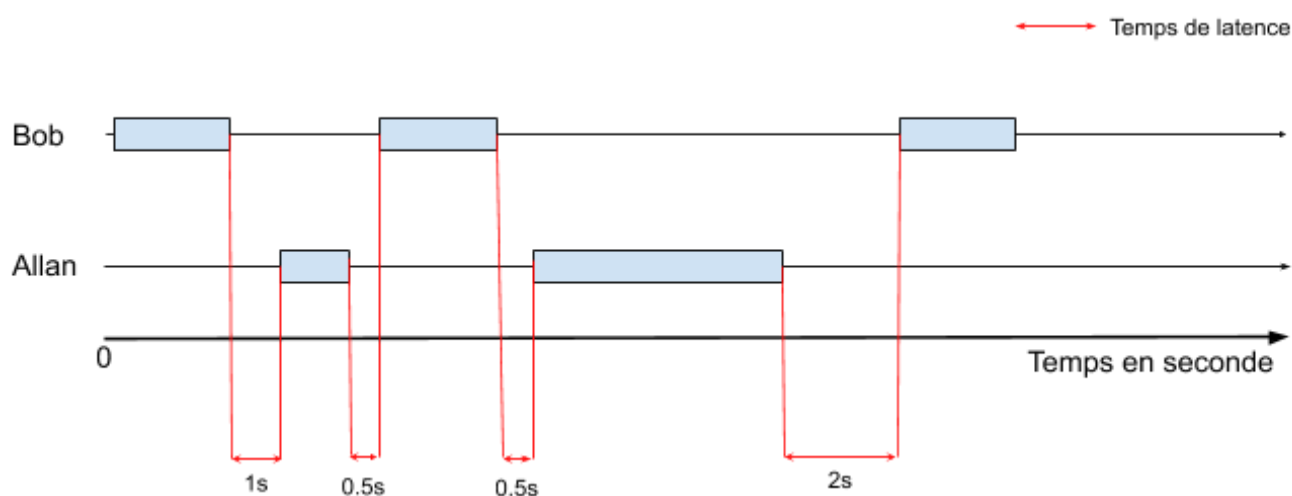


Figure 9: Exemple de conversation entre Bob et Allan avec les temps de latences.

## 2) Les différentes parties du calcul de la synchronie interpersonnelle

### a) Le taux de latence

La première partie du calcul de la synchronie interpersonnelle est l'évolution du taux de latence lors d'un échange, d'abord entre deux personnes, puis entre plusieurs personnes dans un groupe. Le problème de l'application de cette méthode à un groupe est que pour chaque prise de parole d'une personne, elle calcule le temps de réponse de la deuxième personne impliquée. Sauf que dans un groupe, il y a rarement tout le monde qui répond à la personne qui parle. On ne doit donc pas prendre en compte les personnes qui ne répondent pas. Ce problème créerait régulièrement des temps de latence longs et fausserait donc le résultat. Par exemple, sur la figure 10, on peut voir qu'il est inutile de calculer le temps de latence entre Charles et Allan puisque Allan ne répond pas à Charles mais à Bob.

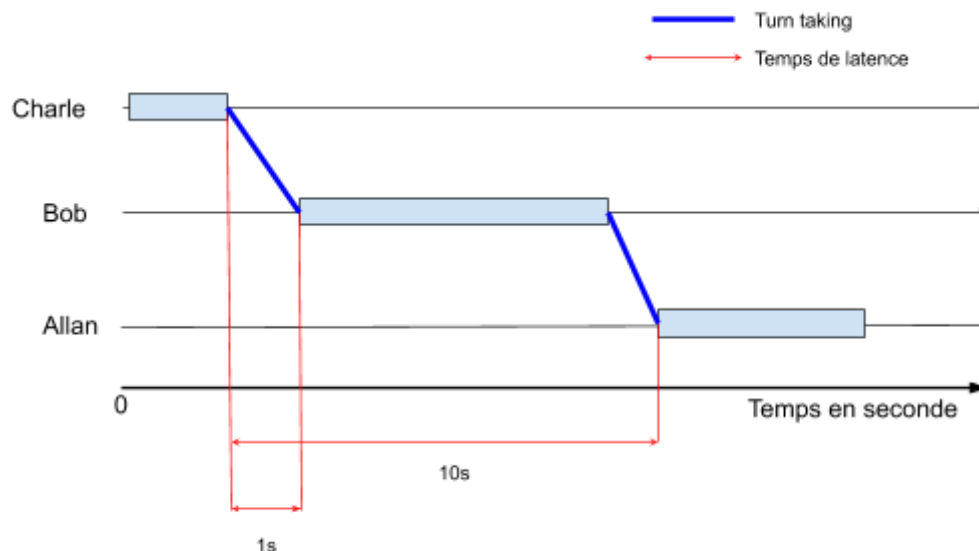


Figure 10: Temps de latence et turn taking entre Charles, Bob et Allan

La méthode utilisée pour calculer le taux de latence globale récupère, lorsqu'une personne parle, le temps de réponse minimal chez tous les autres membres du groupe. Une fois tous les temps de réponse calculés, l'algorithme calcule le taux de latence pour chaque personne et retourne la moyenne des ratios de synchronie de chaque personne.

Cette méthode permet de savoir qu'elle est la personne qui répond à l'orateur, et donc, grâce à cela, d'éviter de prendre en compte les temps de latence d'une personne ne répondant pas à l'orateur mais à quelqu'un d'autre, plus tard dans l'interaction. De plus, en faisant des taux de latences, on obtient une vue d'ensemble du taux de latence dans l'interaction.

Pour tester cet algorithme, on a comparé la mesure du taux de latence de la première session du corpus à la même session, mais soit en décalant les audios, soit en les mélangeant. On a donc généré 1000 pistes audio décalées où chaque personne du groupe commence avec un retard aléatoire. On a aussi 1000 pistes audio mélangées où toutes les frames de 1s sont mis dans un ordre aléatoire. On en a choisi 1000 pour avoir un nombre suffisant de pistes biaisées afin d'avoir un résultat plus précis. Les taux de latences des

pistes biaisées devraient être moins élevées que le taux de latence de la session 1. Les résultats sont présentés en figure 11.

Les résultats montrent que le taux de latence de la session 1 est bien meilleur que celui des pistes audio décalées (plus de 95% des pistes biaisées ont une valeur inférieure). Mais ce n'est pas le cas pour les pistes audio mélangées, ou au contraire le taux de latence est très faible par rapport à celles-ci. Cela est dû au nombre d'échanges entre les participants. Si le nombre d'échanges est élevé, alors le taux de latence du groupe aura tendance à être meilleur. En effet un nombre d'échanges élevé signifie une plus petite latence pour une même fenêtre donnée et donc au final, un résultat plus élevé.

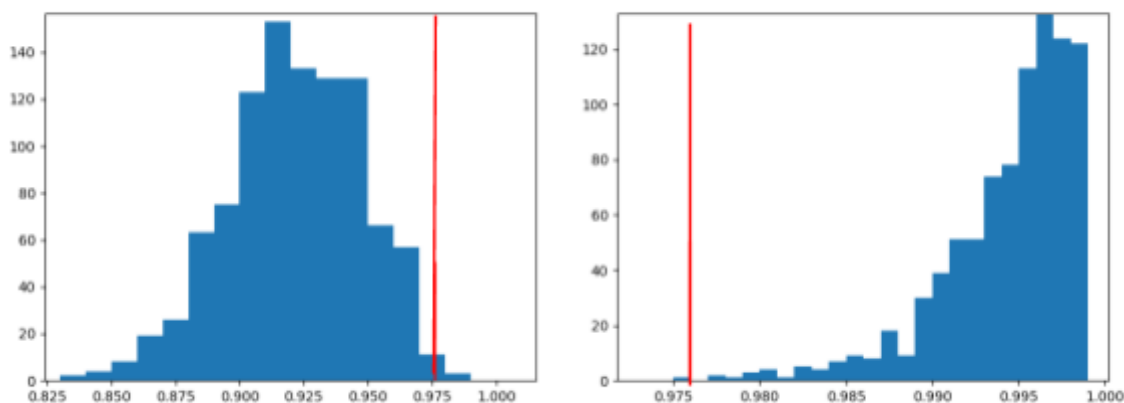


Figure 11: Résultats du taux de latence avec en bleu les pistes audio générées aléatoirement par décalage (figure de gauche) et par mélange (figure de droite) et en rouge la valeur de la session 1.

Le calcul du taux de latence ne se suffisant pas à lui-même au vu des résultats, il est important de prendre d'autres paramètres pour permettre au calcul de la synchronie d'être le plus complet et précis possible.

## b) Le temps de pause

Un des paramètres qui peut-être intéressant lors du calcul du temps de synchronie est le temps de pause. Le temps de pause est important pour la synchronie car moins une personne fait de pauses longues, moins elle hésite sur ce qu'elle va dire, et par conséquent va améliorer la fluidité de l'échange. Ce qui nous intéresse donc pour le temps de pause dans la synchronie est que celle-ci soit faible, et comme il y a plusieurs personnes dans la conversation, on prendra le temps de pause maximal parmi les participants pour obtenir un résultat le plus parlant possible.

Pour tester le temps de pause, on a fait la même chose que pour le taux de latence, ces résultats sont visibles en figure 12. On peut observer sur ces figures que pour les pistes audio générées aléatoirement, le temps de pause est très pertinent puisque si une personne n'arrête pas de parler sans laisser la parole aux autres, forcément le temps de pause s'en trouvera augmenté. Par contre pour les pistes audio générées par décalage, le temps de pause n'indique rien, les participants s'ils parlent normalement auront toujours un temps de pause équivalent, le décalage n'induit donc pas de différence par rapport à la session 1.

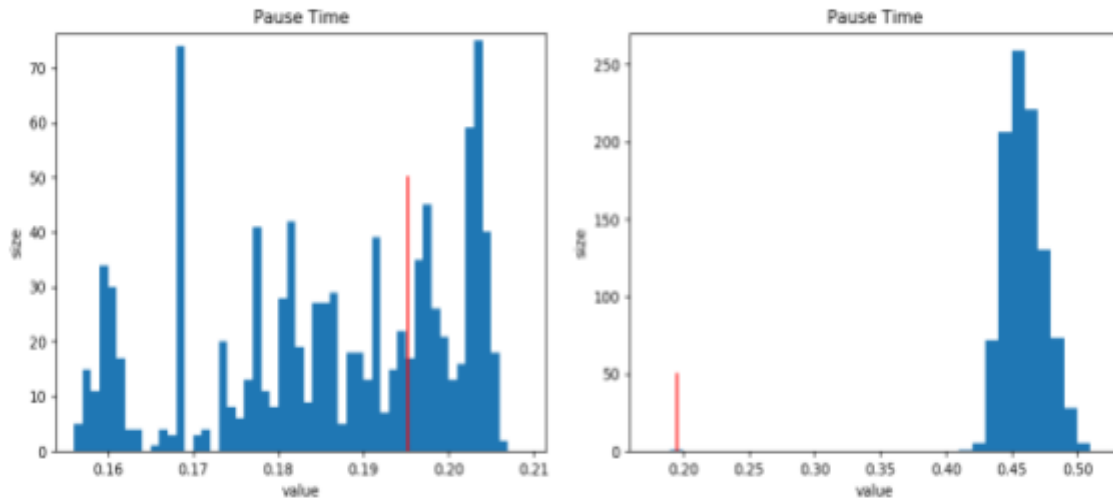


Figure 12: Résultats du temps de pause avec en bleu les pistes audio générées aléatoirement par décalage (figure de gauche) et par mélange (figure de droite) et en rouge la valeur de la session 1.

Le temps de pause n'est pas non plus une valeur qu'on peut utiliser seul, il faut pouvoir l'augmenter avec d'autres paramètres pour rendre encore la mesure de la synchronie précise.

### c) Le temps de parole

Un autre paramètre intéressant dans le calcul de la synchronie interpersonnelle est la somme des temps de parole. En effet, plus l'ensemble des personnes du groupe parle sur une fenêtre donnée, plus ils vont avoir une bonne dynamique. Pour avoir une valeur intéressante de la somme des temps de parole, il faut aussi considérer que si tout le monde parle tout le temps, on ne peut pas avoir une bonne synchronie car personne ne s'écoute, de même que si personne ne parle. Pour pallier ce problème, le calcul du temps de parole comporte un "meilleur résultat" fixé à un sur le nombre de participants pour indiquer qu'une seule personne parle à la fois. Ainsi, plus la valeur de la somme des temps de parole se rapproche de ce résultat, meilleure est la synchronie.

Les tests sur le temps de parole sont les mêmes que sur le taux de latence et le temps de pause, et les résultats sont présentés en figure 13. Ces résultats montrent comme pour le temps de pause que celui-ci n'est pas représentatif lorsqu'on décale les audios de chaque personne du groupe. Mais la valeur du temps de parole de la session 1 reste meilleur que les pistes audio mélangées. Ces résultats sont obtenus car le temps de parole pour une personne est identique quelque soit le moment où cette personne commence à parler. Si on somme le temps de parole des différentes personnes du groupe même en les décalant, on aura ainsi à peu près toujours les mêmes résultats. Alors que pour les pistes audios mélangées, les participants se retrouvent à parler beaucoup et donc à dépasser la barre du "meilleur résultat", ce qui fait que la valeurs du temps de parole est plus élevée que celle de la session 1.

Le temps de parole est au final une assez bonne valeur pour évaluer la synchronie et elle permet donc d'améliorer le résultat final, mais il y a encore une dernière composante qui peut être intéressante, le silence.

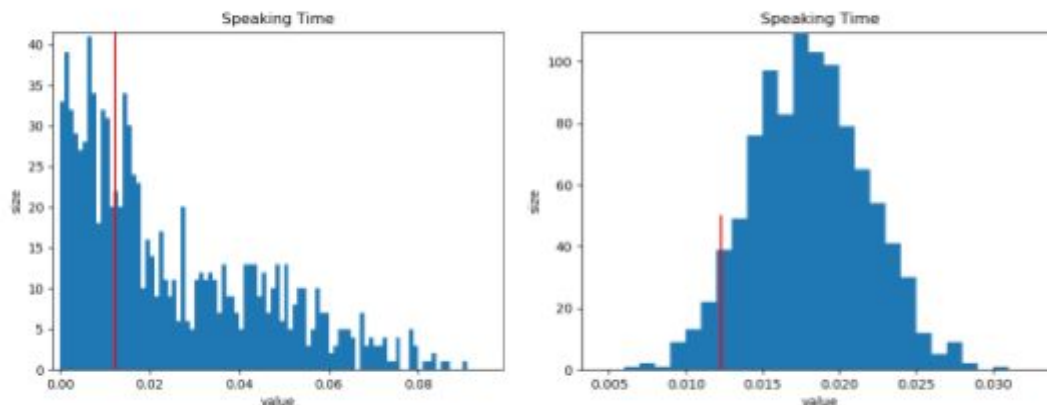


Figure 13: Résultats du temps de parole avec en bleu, les pistes audio générées aléatoirement par décalage (figure de gauche) et par mélange (figure de droite) et en rouge la valeur de la session 1.

#### d) Le silence

Le dernier paramètre dans le calcul de la synchronie interpersonnelle est le silence. Pour la synchronie, il est important qu'il n'y ait pas beaucoup de silence, car les silences montrent qu'il n'y a personne qui veut reprendre la parole. Mais il ne faut pas qu'il y en ait trop car dans ce cas, comme pour le temps de parole, cela peut indiquer que tout le monde parle en même temps, ce qui fait que la synchronie serait basse. Pour évaluer une bonne quantité de silence, il faut un seuil qui est dans l'algorithme le temps de pause. Ce seuil est fixé ainsi car si une personne prend la parole pendant que quelqu'un d'autre fait une pause cela peut être considéré comme une interruption non voulue, ce qui n'aide pas à la coordination des différentes personnes du groupe, donc à la synchronie.

Les tests sont toujours les mêmes et sont présentés dans la figure 14. On peut observer que pour le silence, que ce soit par décalage ou par mélange, la valeur du silence est meilleure que les pistes biaisées. Les tests étant aléatoires, il n'est pas rare qu'il y ait quelqu'un qui parle tout le temps, ce qui entraîne une valeur de silence très basse et ainsi une grande différence entre la valeur de pause et la valeur de silence. C'est ce qui fait que les résultats sont meilleurs.

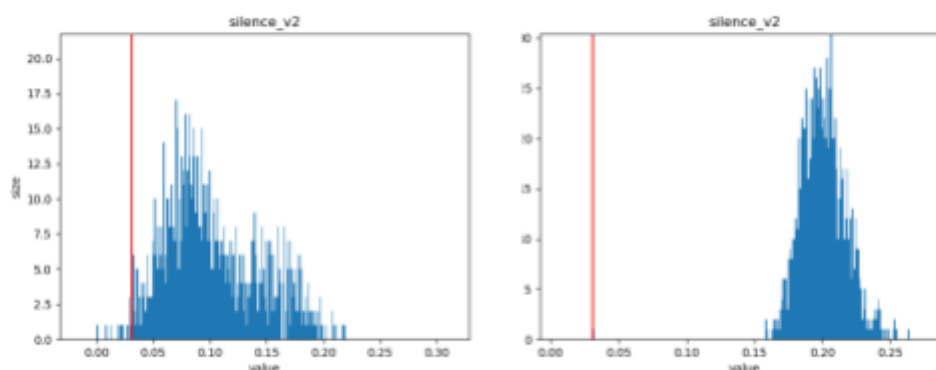




Figure 14: Résultats du silence avec en bleu les pistes audio générées aléatoirement par décalage (figure de gauche) et par mélange (figure de droite) et en rouge la valeur de la session 1.

### e) Le calcul final de la synchronie interpersonnelle

Après avoir obtenu les quatre paramètres de la synchronie: le ratio de synchronie du groupe, le temps de pause, le temps de parole total et le silence, il faut pouvoir mixer ces quatre résultats pour obtenir une valeur de synchronie. La première étape de ce calcul est de normaliser chacune des valeurs entre 0 et 1, puis de décréter si la valeur pour avoir une bonne synchronie est 0 ou 1. Ainsi, tous les paramètres expriment une meilleure synchronie avec une valeur de 0 plutôt que de 1 sauf pour le taux de latence qui exprime une meilleure synchronie pour une valeur de 1. Avec ceci, il nous reste à faire une moyenne de ces valeurs.

On peut observer les résultats des tests habituels sur la figure 15. On peut ainsi voir que le calcul de la synchronie est très efficace sur les deux tests et qu'on a plus de 98% des pistes audio générées aléatoirement qui ont une moins bonne synchronie. Les pistes mélangées sont bien moins bonnes que les pistes décalées. Ceci s'explique par le fait que dans une conversation entièrement mélangée, il n'y a même plus de logique au niveau de l'individu, alors qu'il reste cette logique pour les pistes audio décalées puisque la piste audio reste intacte.

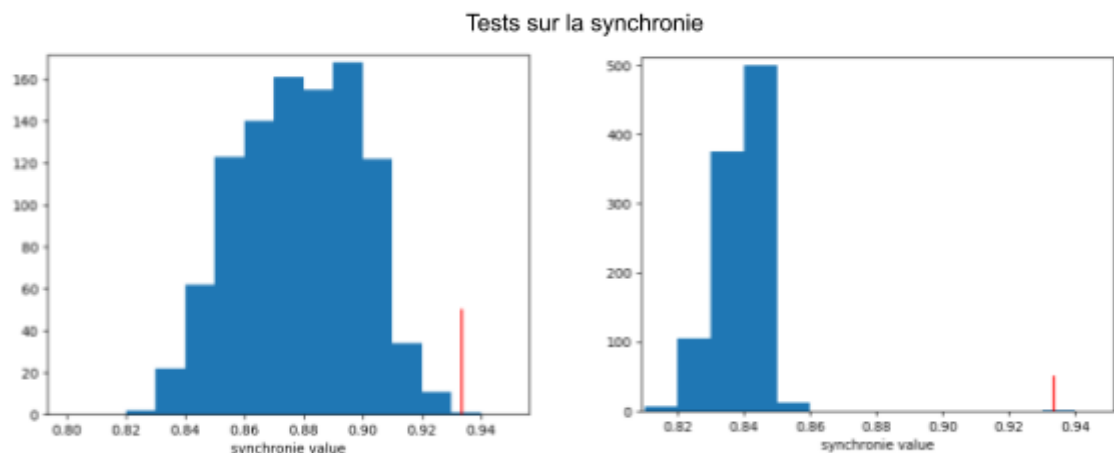


Figure 15: Résultats de la synchronie avec en bleu les pistes audio générées aléatoirement par décalage (figure de gauche) et par mélange (figure de droite) et en rouge la valeur de la session 1.

## 3) Les tests

Les tests précédents permettent de montrer que le calcul de la synchronie fournit un meilleur résultat qu'un ensemble de résultats biaisés et ceci seulement sur la session 1, ce n'est donc pas suffisant. Il faut donc pouvoir tester ce résultat, premièrement sur un ensemble de vidéos différentes, deuxièmement en présence d'autres personnes que l'on sait moins bonnes, et troisièmement, il va falloir vérifier si ce qu'on calcule est bien un résultat cohérent et correspond bien à la synchronie.

### a) La synchronie sur différentes sessions

On peut observer les résultats de l'évolution de la synchronie au cours de différentes sessions tirées de notre corpus sur la figure 16. La fenêtre de 2 minutes a été choisie puisqu'elle est suffisamment large pour avoir assez d'échanges et d'avoir une valeur de ratio de synchronie plus précise. Elle est aussi suffisamment petite pour avoir une quantité suffisante de données sur l'ensemble de l'échange, et aussi parce que la cohésion a été annotée toutes les 2 minutes.

On peut voir que la synchronie au cours du temps fluctue, mais aussi que sur chacun des graphes, les valeurs de synchronie restent principalement entre 0.80 et 0.90. On peut en déduire que la synchronie entre les différentes sessions est stable, l'algorithme donne des valeurs identiques sur plusieurs sessions, et il n'est ainsi pas spécifique. Après observation de la décomposition des différents résultats et de leurs variations, on peut déduire que la présence d'une faible synchronie est due au nombre de passation de paroles : si une seule personne parle durant la fenêtre de 2 minutes, il sera difficile de déterminer la synchronie, ou bien à la très faible intervention d'une personne du groupe dans la fenêtre de calcul, alors qu'une haute synchronie est plutôt synonyme de la présence de toutes les personnes et d'un échange régulier de paroles.

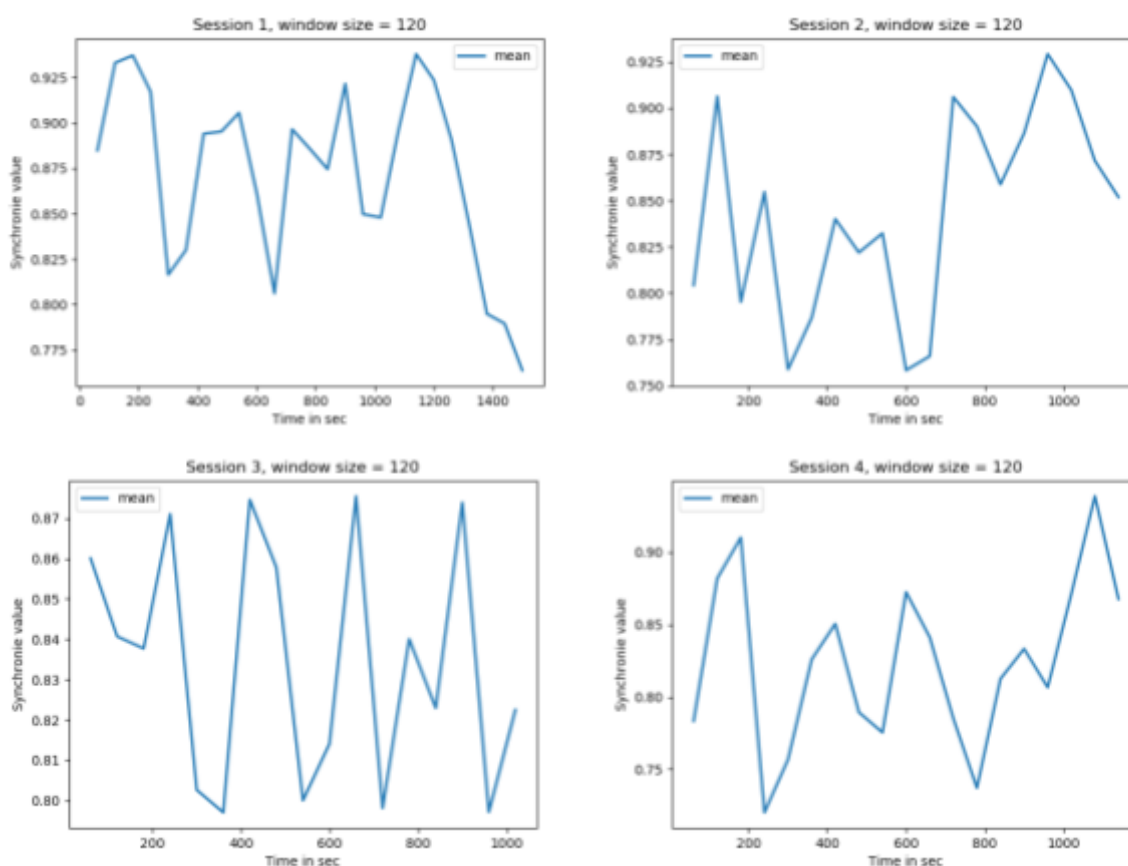


Figure 16: Valeur de synchronie sur 4 sessions différentes, calculée toutes les 60s sur une fenêtre de temps de 2 minutes.

## b) La synchronie en cas d'ajout ou suppression de personne

Sur la figure 17, on peut observer les variations de la synchronie en fonction de si on ajoute, remplace ou enlève une personne. On peut ainsi remarquer que la synchronie de la session 1 n'est pas toujours la meilleure. Par exemple dans la figure en haut à gauche, on peut voir que la fin de la courbe avec deux personnes passe au-dessus de la courbe test. Ceci est dû au fait que la troisième personne dans la session 1 fait descendre la synchronie et donc en l'enlevant, la courbe remonte. On peut aussi voir que quand on ajoute une personne (figure en haut à droite), la courbe suit de très près la courbe test, parce que la synchronie des trois premières personnes reste présente et que la dernière personne va plus ou moins changer ces valeurs-là.

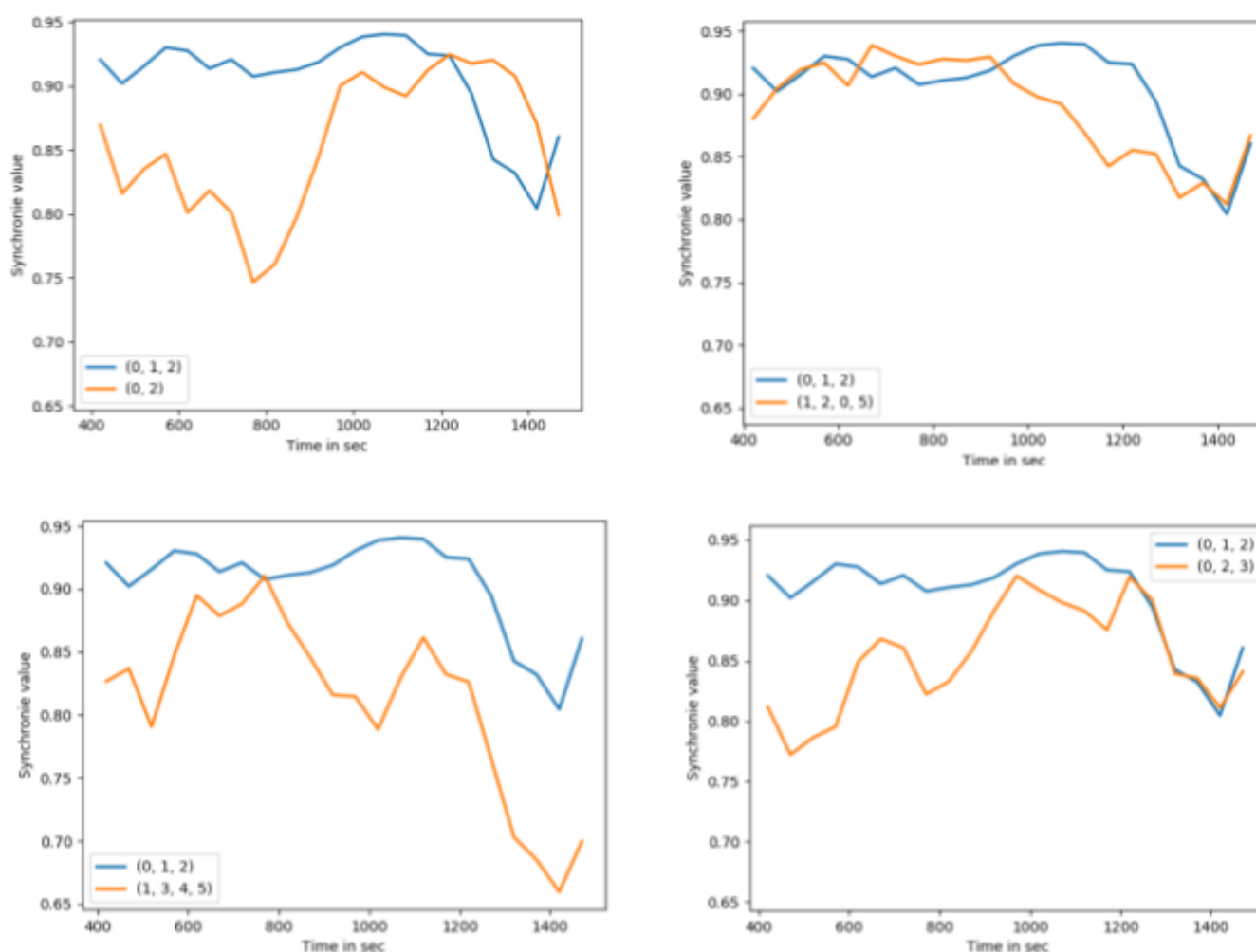


Figure 17: Les différentes valeurs de synchronie entre la session 1 (en bleu) et en enlevant une personne(en haut à gauche), en ajoutant une personne aléatoire(en haut à droite), en mettant 4 personnes aléatoires(en bas à gauche) et en changeant une personne(en bas à droite)

### c) Le test du calcul de la synchronie

Le dernier test à effectuer sur l'algorithme du calcul de synchronie interpersonnelle de groupe est de voir si le calcul effectué représente une bonne approximation de la synchronie, et que la mesure ne soit pas juste aléatoire. Pour ceci, toutes les fenêtres de deux minutes des sessions du corpus ont été annotées sur une série de 10 questions (voir figure 18) pour déterminer la cohésion. La cohésion selon Carron (1982) est "un processus dynamique qui se caractérise par la tendance d'un groupe à se serrer les coudes et à demeurer unis dans la poursuite de ses objectifs". Les premières questions représentent la cohésion de tâche, une cohésion centrée sur les actions des personnes présentes dans le groupe et sur l'objectif de celui-ci. Les 5 dernières questions sont sur la cohésion sociale qui est basée sur les aspects de communications et d'interactions entre les différentes parties du groupe. La cohésion est une valeur proche de la synchronie, car pour avoir bonne coordination temporelle, il faut en entre autre savoir quelle est l'objectif du groupe et savoir s'aider au sein du groupe, il faut donc entre autre une bonne cohésion. Et inversement pour avoir une bonne cohésion, il faut entre autre avoir une bonne synchronie. L'objectif de ce test est donc de voir si la cohésion annotée et la synchronie calculée ont une bonne corrélation. Si c'est le cas, on pourra dire que la synchronie correspond à la mesure voulue. Vu que l'on a trois types de cohésions différentes : la cohésion de tâche (questions 1 à 5), la cohésion sociale (questions 6 à 10) et la cohésion générale (questions 1 à 10), on a calculé les corrélations entre chacun de ces types de cohésion et la synchronie, la cohésion étant calculée en faisant une moyenne entre les différentes questions.

On peut observer sur la table 2 les différentes corrélations. Pour les cinq premières questions on peut voir que la synchronie est plutôt bonne pour les trois premières sessions et moins bonnes pour la quatrième. Pour les cinq dernières sessions, la corrélation est entre 0.38 et 0.75 qui sont des valeurs raisonnables. Quand on prend les dix questions ensemble, on peut observer que la corrélation moyenne est de 0.63. On peut en déduire que la corrélation est en général assez bonne quel que soit le type de cohésion, avec quelques exceptions comme pour la session 4, dans les cinq premières questions, et la session 2, dans les 5 dernières. On pourrait dire que la synchronie représente bien ce que l'on veut, mais on a un calcul de la cohésion assez simple qui consiste en une moyenne, alors que chaque question doit avoir un poids différent pour exprimer son importance.

Session	1	2	3	4
Corrélation questions 1-5	0.78	0.57	0.68	0.36
Corrélation questions 6-10	0.75	0.38	0.68	0.60
Corrélation questions 1-10	0.79	0.51	0.70	0.55

Table 2: Corrélation entre la cohésion et la synchronie en fonction de la session et des différentes questions impliquées.

Pour mieux vérifier que le calcul de synchronie donne bien la bonne valeur, on peut déterminer à partir des 10 questions quelles sont celles qui sont importantes dans le calcul de la synchronie, et celles qui au contraire ne le sont pas. On peut ainsi améliorer notre corrélation entre les différents types de cohésion et la synchronie. Pour cela, on a utilisé un modèle mixte linéaire pour donner à chaque question un poids. Plus ce poids se rapproche de 0 moins la question va être importante dans la synchronie, et inversement plus le poids est éloigné de 0 plus la question sera importante pour la synchronie. L'ensemble des questions est dans la figure 18, et les poids aux questions dans la table 3.

1. Chaque membre de l'équipe semble avoir suffisamment de temps pour apporter sa contribution.
2. L'équipe semble partager la responsabilité de la tâche.
3. Dans l'ensemble, les membres de l'équipe semblent collaborer.
4. J'ai le sentiment que les membres de l'équipe partagent les mêmes buts/objectifs/intentions.
5. Dans l'ensemble, les membres du groupe donnent beaucoup de feedback.
6. Dans l'ensemble, les membres du groupe s'écoutent attentivement.
7. Dans l'ensemble, les membres de l'équipe semblent se soutenir les uns les autres.
8. Dans l'ensemble, le groupe semble être à l'écoute et synchronisé les uns avec les autres.
9. Dans l'ensemble, je pense que le groupe de travail fonctionne spontanément.
10. Dans l'ensemble, les participants semblent être impliqués/engagés dans la discussion.

Figure 18: Les 10 questions du questionnaires pour déterminer la cohérence.

Question:	1	2	3	4	5	6	7	8	9	10
Poids en séparants les deux séries	0.015	-0.009	0.022	-0.001	0.001	0.031	0.004	-0.006	-0.013	0.016
Poids en gardant toutes les questions	0.008	-0.001	0.013	0.013	-0.010	0.033	-0.008	-0.022	-0.006	0.012

Table 3: Les poids des questions selon un modèle mixte linéaire. En rouge les questions les moins importantes pour la synchronie et en vert les questions les plus importante pour la synchronie.

On peut observer dans les résultats, tout d'abord en séparant les deux séries, que les questions les plus importantes sont les numéros 3 et 6. La question 6 est un élément de fluidité important, car si dans le groupe les personnes s'écoutent, alors la conversation est plus fluide. La question 6 vient aussi en première place quand on prend les 10 questions, la question 3 met en lumière la collaboration et donc une dynamique de groupe. Les questions les moins importantes sont les numéros 4, 5 et 7 en séparant les deux séries et 2, 9 en prenant toutes les questions. Ce sont toutes des questions qui n'ont pas d'impact direct sur

la synchronie, ce qui est donc normal qu'elles soit peu utilisées. La deuxième question la plus importante quand on prend toute la série est la numéro 8, ce qui est en accord avec la définition de la synchronie, puisque la question parle de synchronisation et d'écoute. Les nouvelles corrélations calculées avec les poids sont montrées dans la table 4 avec les erreurs standards obtenues, qui elles sont dans la table 5.

Session	1	2	3	4
Corrélation questions 1-5	0.87	0.45	0.75	0.41
Corrélation questions 6-10	0.79	0.42	0.76	0.84
Corrélation questions 1-10	0.86	0.53	0.70	0.84

Table 4: Corrélations entre la synchronie et la cohésion en utilisant les poids calculés par le modèle mixte linéaire.

Question:	1	2	3	4	5	6	7	8	9	10
Erreurs standard en séparant les deux séries	0.10	0.014	0.015	0.014	0.013	0.011	0.011	0.014	0.010	0.009
Erreurs standard en gardant toutes les questions	0.09	0.013	0.016	0.017	0.015	0.012	0.014	0.018	0.012	0.012

Table 5: Erreurs obtenues par le modèle mixte linéaire.

Les nouvelles corrélations sont en général meilleures, voir bien meilleures, augmentant au maximum de 0.30 pour la session 4. Mais ce n'est pas le cas pour la session 2 qui voit toutes ses corrélations diminuer, d'au moins 0.10. Pour les erreurs standard, on peut déjà remarquer que la différence entre elles, en séparant les deux types de cohésion ou en prenant la cohésion générale, reste à peu près identique. Le maximum de différence obtenue est 0.004. L'erreur standard en fonction des questions oscille entre 0.009 et 0.018, ce qui n'est pas énorme. On peut ainsi en conclure avec ces résultats-là que le modèle mixte linéaire marche bien et que les résultats donnés par celui-ci sont corrects. On peut ainsi accepter les poids donnés plus haut. En conclusion, le calcul de la synchronie donne la valeur voulue, mais celui-ci n'est pas parfait : les résultats sur les poids associés aux questions le montrent bien, comme la différence entre les cohésions des sessions 1, 3 et 4 par rapport à la session 2.

## VI) Conclusion et suite

La synchronie interpersonnelle est une valeur importante pour déterminer si un groupe arrive à communiquer de manière fluide et dynamique. La synchronie décrite ici est ainsi découpée en quatre parties: le taux de latence, le temps de pause, le temps de parole et le silence, chacun ajoutant et améliorant la précision du calcul. L'ensemble des tests effectués montre que la synchronie obtenue est indépendante vis à vis du nombre de personnes et des vidéos utilisées, et que le calcul correspond à la synchronie. Mais il pourrait y avoir encore des améliorations vis-à-vis de son calcul, par exemple avec l'ajout d'autres paramètres plus complexes en vocal, mais aussi l'ajout de paramètres non verbaux. En effet, le non verbal joue au moins autant sur la synchronie que le verbal. On pourrait aussi calculer des poids à chacun des paramètres, il est peu probable que chaque paramètre ait la même importance dans la synchronie interpersonnelle. Il faudrait aussi pouvoir tester les résultats sur d'autres corpus, pour observer si les résultats restent constants et ne se dégradent pas dans d'autres types de groupes.

## Bibliographie

- Ambady, Nalini, et Robert Rosenthal. 1992. « Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis ». *Psychological Bulletin* 111 (2): 256-74. <https://doi.org/10.1037/0033-2909.111.2.256>.
- Burgoon, Judee K., Laura K. Guerrero, et Kory Floyd. 2016. *Nonverbal Communication*. Routledge.
- Delaherche, Emilie, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, et David Cohen. 2012. « Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines ». *IEEE Transactions on Affective Computing* 3 (3): 349-65. <https://doi.org/10.1109/T-AFFC.2012.12>.
- Hung, Hayley, et Daniel Gatica-Perez. 2010. « Estimating Cohesion in Small Groups Using Audio-Visual Nonverbal Behavior ». *Multimedia, IEEE Transactions on* 12 (novembre): 563-75. <https://doi.org/10.1109/TMM.2010.2055233>.
- Ruede, Robin, Markus Müller, Sebastian Stüker, et Alex Waibel. 2017. « Yeah, Right, Uh-Huh: A Deep Learning Backchannel Predictor ». *arXiv:1706.01340 [cs]*, juin. <http://arxiv.org/abs/1706.01340>.
- Shaw, Marvin E. 1971. *Group Dynamics, the Psychology of Small Group Behavior*. McGraw-Hill.
- Skantze, Gabriel, Martin Johansson, et Jonas Beskow. 2015. « Exploring Turn-taking Cues in Multi-party Human-robot Discussions About Objects ». In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 67–74. ICMI '15. New York, NY, USA: ACM. <https://doi.org/10.1145/2818346.2820749>.
- Varni, Giovanna, Marie Avril, Adem Usta, et Mohamed Chetouani. 2015. « SyncPy: A Unified Open-source Analytic Library for Synchrony ». In *Proceedings of the 1st Workshop on Modeling INTERPERSONAL Synchrony And influence*, 41–47. INTERPERSONAL '15. New York, NY, USA: ACM. <https://doi.org/10.1145/2823513.2823520>.