



# AGORA- PHONE

Humanisation du numérique  
par la parole échangée

**Mémoire d'Andy Bruneteau**

Étudiant  
au Mastère spécialisé Innovation by Design  
de l'ENSCI - Les Ateliers

**Réalisé sous la direction de Roland Cahen**

Enseignant chercheur,  
en charge du studio sonore de l'ENSCI,  
membre de l'EA du Centre de Recherche  
en Design ENSCI - ENS Paris-Saclay



## Remerciements

---

La réalisation et l'écriture de ce mémoire n'auraient pu voir le jour sans la contribution de près ou de loin de plusieurs personnes que je souhaite, à travers ces lignes, remercier.

Je remercie mon directeur de mémoire Roland Cahen pour m'avoir partagé son expertise du design, ses connaissances et sa méthodologie de travail ; pour son encadrement et d'avoir pris la peine de me guider dans mes heures de doute, ainsi que parmi mes pensées nébuleuses. L'appui apporté et le temps consacré ont été considérables, je lui en suis reconnaissant.

Un grand merci à toute l'équipe pédagogique du mastère spécialisé Innovation by Design : Mathias Béjean, Stéphane Gauthier, Geneviève Sengissen et Muriel Zerafa pour l'encadrement et la qualité de la formation. Un remerciement tout particulier aux deux responsables pédagogiques du mastère, Mathias Béjean et Stéphane Gauthier, pour leurs enseignements, leur passion de l'innovation, l'énergie apportée tout au long de l'année et la qualité des intervenants que vous nous proposez. Je remercie l'ensemble des intervenants pour la richesse des enseignements et de nous ouvrir les yeux pour apprendre à regarder le monde différemment.

Un merci à l'ensemble des employés de l'EN-SCi-Les Ateliers, ainsi qu'aux étudiants de la formation initiale qui partagent sans cesse leurs créations, source d'inspiration.

Je remercie toute l'équipe de Meaningful : Marine Rouit-Leduc, Nicolas Rouit-Leduc, Galilée Al-Rifaï, Antoine Martinez, Jennifer Tytgat, Camille Jacoupy et Juliette Magniere, de m'avoir partagé leur vision du design, du monde industriel et encouragé dans mon travail au cours des mois passés avec eux.

Un grand merci à l'ensemble de mes collègues du mastère spécialisé Innovation by Design pour cette année pleine d'émotions et de souvenirs. C'est vous, avant tout, qui êtes la richesse de la formation. Merci pour vos contributions, vos partages d'expériences et visions du monde, ainsi que pour la solidarité dont vous avez fait preuve dans cette traversée commune.

À ma famille, pour le soutien et les encouragements qui ont été permanents depuis mon retour sur les bancs d'école.

Et non des moindres, je remercie affectueusement Garance Plissonnier pour son soutien infaillible, ses encouragements constants, ses sacrifices, sa patience et ses contributions tout au long de plus d'un an de formation. Cette réussite n'est pas que mienne !

Un grand Merci !



Sommaire

---

Remerciements (p.7)

Introduction (p.15)

## **Make the world a better place**

Chapitre 01 (p.23)

### **La vive voix**

Chapitre 02 (p.37)

### **La parole dans les technologies**

Chapitre 03 (p.51)

# **Comment pourrions-nous ?**

Conclusion (p.75)

Bibliographie (p.79)

Annexes (p.83)

# Make the world a better place

Depuis plus de deux décennies, de nombreuses entreprises du numérique œuvrent à simplifier notre vie en résolvant des problématiques du quotidien. Ces entreprises exploitent la dématérialisation, qui nous libère de certaines contraintes physiques, pour nous permettre de réaliser quantité d'activités en toute autonomie. Ainsi, il y a encore quelques années, qui aurait pu imaginer pouvoir ouvrir un

compte bancaire et réaliser ses opérations sans l'intervention d'un banquier<sup>1</sup> ? ; visiter des biens immobiliers sans l'accompagnement d'un agent<sup>2</sup> ? ; Louer une voiture sans passer par le comptoir de réservation<sup>3</sup> ? ; Ordonner à son assistant virtuel de prendre rendez-vous chez le coiffeur<sup>4</sup> ? La technologie numérique nous permet désormais de réaliser toutes ces activités. Toutefois, devant toutes ces innovations me vient une question :

---

<sup>1</sup> La banque N26

<sup>2</sup> Opendoor

---

<sup>3</sup> Virtuo

<sup>4</sup> Google Duplex

# où sont donc passés les échanges humains ?

Un premier élément de réponse m'est apparu lorsque, alors chef de projet innovation, j'ai conçu un service numérique pour le compte d'une grande entreprise. Durant quelques mois, notre équipe de projet a étudié les opportunités de la transformation digitale sur l'un des métiers de la promotion immobilière : le développement foncier. La démarche a abouti sur le développement d'un service numérique mettant en relation des propriétaires immobiliers avec des promoteurs. Lors de la conception du parcours utilisateur nous avons fait le choix de restreindre les échanges directs avec nos experts immobiliers aux propriétaires ayant validé certaines étapes du parcours. Ainsi, une forte majorité des propriétaires inscrits recevaient de notre entreprise uniquement des mails automatisés et standardisés. Très coûteux, les échanges humains ont donc été réduits au minimum pour permettre à notre jeune entreprise d'atteindre ses objectifs de performance économique. Pour autant, la conception d'un parcours linéaire nous permet de créer l'illusion d'interactions sociales qui sont réalisées par la machine. Ce parti pris n'est pas un cas isolé : la nécessité de répondre aux enjeux d'économies d'échelles et d'industrialisation pousse les entreprises numériques à automatiser les interactions permettant ainsi de reproduire invariablement et à l'infini un parcours. In fine, les échanges humains se retrouvent bannis progressivement car plus complexes.

Cette expérience dans l'entrepreneuriat me fait constater une intermédiation de la technologie numérique dans les relations humaines avec pour résultat de nous isoler de nos pairs. Toutefois, les expériences proposées par les start-up ne sont qu'une infime partie de notre exploitation de cette technologie. Qu'en est-il alors du reste de notre utilisation du numérique ?

Ce sujet, Sherry Turkle l'étudie depuis plus de 30 ans dans le cadre de ses recherches sur les effets sociaux et psychologiques des nouvelles technologies qu'elle réalise au sein du Massachusetts Institute of Technology (MIT). Dans son dernier ouvrage, *Seuls Ensemble* [Sherry Turkle, 2011], l'auteur y présente ses recherches sur l'état actuel de notre société où les technologies nous permettent d'entretenir des relations à distance de notre intimité et de notre authenticité. Selon la professeure, cette volonté de communiquer indirectement est née d'un changement de perception des relations humaines désormais vécues comme difficiles, frustrantes et risquées émotionnellement. Pour se protéger et préserver son intimité,

l'être humain se réfugie derrière la technologie et l'impose comme intermédiaire dans les relations humaines afin d'éviter les échanges directs, jugés trop complexes. Une des conséquences de cette situation est que l'être humain connaît désormais bien moins ses proches. Ceci est notamment exposé par Sherry Turkle avec ce propos recueilli dans le cadre de ses recherches :

**« Je préfère les textos, Twitter, le mur Facebook d'un ami. J'y trouve tout ce que j'ai besoin de savoir. »**

Figure i.1. Illustration  
Photographie issue du projet « Screen Time » de Dafydd Jones et reflet du travail de Sherry Turkle sur la place de technologie dans nos vies.



Finalement, quelques universitaires mettent en évidence que l'intermédiation de la technologie numérique dans les relations humaines ne se restreint pas aux domaines d'activités des start-up. Celui-ci est généralisé et porte une tendance qui nous mène vers davantage d'interactions avec la machine au détriment des échanges humains qui s'amenuisent, à l'image du film *Her* qui projette l'idée de la possible relation amoureuse avec la machine (fig i.2). Dans ce mémoire, je souhaite donc questionner cette intermédiation de la technologie numérique en recherchant des pistes de solutions dont la finalité est une communication entre pairs plus directe et plus humaine.

Un des axes de travail évidents semble être le retour des échanges humains dans la vie réelle et physique. Toutefois, avec une technologie nu-

mérique désormais bien ancrée dans les usages de communication, je ne crois nullement que l'être humain va s'en priver. L'idée de ce mémoire n'est pas de mettre de côté la technologie numérique, mais à l'inverse d'explorer le vaste champ des possibles qu'offre la dématérialisation. D'autre part, au fil de mon parcours, j'ai acquis la conviction que la technologie se doit être au service de l'humain, de la société et non être un intermédiaire placé entre chacun de nous. Selon moi, sortir de cette vision techno-centrée peut être source de nouvelles innovations. Pour autant, les axes de travail restent nombreux et, quelque peu démunis face à cette infinité, ce sont les deux fondateurs de l'agence de design Les Sismo qui vont me souffler l'élément central de ma réponse : la parole humaine.

Figure i.2. Illustration

*Her* de Spike Jonze (2013) met en scène Joachim Phoenix dont le personnage vit une histoire d'amour avec Samantha, une intelligence artificielle.

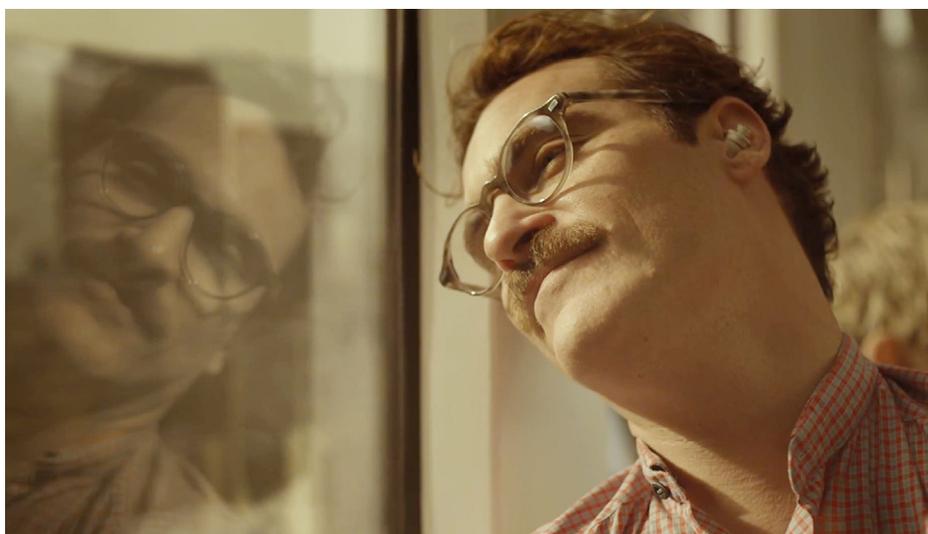


Figure i.3. Illustration

Une des pages du site web de l'agence de design Les Sismo. L'interface graphique est accompagnée d'un lecteur audio ancré dans le bas de la page. Sur celui-ci est indiqué : « Découvrez les Sismo en écoutant ses fondateurs en parler »

C'est au détour d'une visite de la page web Les Sismo (fig. i.3) que je constate la présence d'un lecteur audio ancré dans le bas de la page. Celui-ci donne accès à plusieurs enregistrements de présentation de l'agence effectués par les deux fondateurs, Antoine Fenoglio et Frédéric Lecourt. Je suis alors saisi par la complémentarité et la diversité d'informations présentes sur l'interface graphique et dans les enregistrements audio.

D'un côté, l'interface graphique, objet statique et inerte, est composée de blocs de texte proposant un discours formel et commercial qui fait écho aux besoins du marché. Le contenu semble parfois impersonnel et ne délivre que les informations essentielles dans un but unique : transmettre les données importantes.

À l'opposé, les enregistrements audio sont moins avarés de détails et nécessitent d'y consacrer quelques minutes. La parole est vivante, pleine de nuances et riche d'informations contextuelles. Ces enregistrements nous font aussi quitter la rigueur du discours de vente pour laisser place à plus de spontanéité et aux anecdotes de l'entreprise. L'agence commence à prendre vie, à devenir humaine et on y entre par ces objets sonores qui nous révèlent les personnalités de celles de ceux qui y travaillent. C'est une véritable rencontre avec Antoine Fenoglio et Frédéric Lecourt que nous avons et que l'on souhaiterait transformer en conversation. La parole humaine semble donc avoir un avantage incontestable dans la richesse d'informations qu'elle délivre et le sentiment de connexion qu'elle crée.

L'utilisation de la parole enregistrée par Les Sismo fait écho à l'instrumentalisation récente de la parole humaine dans le numérique. Il est désormais difficile d'échapper aux assistants vocaux qui s'immiscent dans nos vies privées, nous écoutent et nous parlent. Malgré leurs jeunes âges, les interfaces hommes-machines vocales ont fait l'objet de nombreux fantasmes et ce sont les arguments de facilité d'usage et d'une communication naturelle qui sont désormais au coeur des

discours de vente. Il est vrai que de par sa présence quotidienne dans nos conversations, la parole nous semble innée. Toutefois, cette action ne nous permet-elle pas encore de nous distinguer de la machine ? Ce contexte m'amène à considérer la parole humaine comme un enjeu supplémentaire dans la problématique d'intermédiation où les interfaces hommes-machines vocales représentent un pas supplémentaire. Ainsi, ce mémoire répondra à la problématique suivante :

## la présence de paroles humaines enregistrées dans les interfaces numériques peut-elle favoriser des interactions humaines plus empathiques ?

Pour répondre à la problématique, le premier chapitre fera l'objet de la parole humaine et répondra à ces questions : quelles sont les spécificités et les informations que portent la parole ? Quels sont les bénéfices pour la communication ? Le deuxième chapitre portera un regard sur les usages de la parole humaine dans les technologies actuelles et passées. Ce travail aura pour objectif de mettre en lumière les zones à explorer en ce qui

concerne la parole humaine dans le numérique. Ce mémoire se conclura sur la conception d'un cas réel permettant de sortir de l'abstraction, de proposer une perspective de projet innovant et d'imaginer de nouveaux usages numériques. Les réponses apportées par la réalisation de ce processus créatif permettront d'agrémenter la discussion posée par la problématique.

Et quand vous parliez, à dessein distrait,  
Je prêtais l'oreille à votre secret :  
Car la voix, ainsi que les yeux de Celle  
Qui vous fait joyeux et triste, décèle,  
Malgré tout effort morose et rieur,  
Et met au plein jour l'être intérieur.

Paul Verlaine

La Bonne Chanson, XIII, 1870

Chapitre 01

---

# La vive voix

C'est avec ces quelques mots de Paul Verlaine, issus du recueil de poème *La Bonne Chanson*, que je souhaite introduire ce premier chapitre. Au cours de celui-ci, nous tenterons de répondre aux questions suivantes : quelles sont les spécificités et les informations que portent la parole ? Quels sont les bénéfices pour la communication ?

En évoquant les émotions et les sentiments, Paul

Verlaine nous ouvre la voie sur les dessous de la parole humaine. Mais pour répondre à ces questions je m'appuierais principalement sur les travaux d'Iván Fónagy, directeur de recherches au CNRS et linguiste de formation, de Joana Révis, chercheuse en linguistique, docteur en science du langage et orthophoniste au CHU de la Timone à Marseille et de Jean Abitbol, ancien chef de clinique à la faculté de médecine de Paris, oto-rhino-laryngologiste et phoniatre.

## 1.1. L'expressivité de la parole humaine

Dans son livre *La vive voix*, Iván Fónagy introduit le sujet de la parole humaine et de la psycho-phonétique, son domaine de recherche, avec ces quelques lignes [Fónagy, 1983] :

**« La vive voix s'oppose en français, comme en d'autres langues, à la lettre morte. [...] L'opposition de ces deux termes indique que la vivacité est inhérente à toute communication vocale. Pour concrétiser quelque peu le terme métaphorique de vivacité, on dit également que la parole est plus expressive ».**

Dans cet extrait, Iván Fónagy place l'expressivité comme l'une des différences majeures entre la communication écrite et vocale. Ainsi, il convient de définir cette caractéristique indissociable de la parole humaine. Pour ce faire, l'auteur s'appuie sur le modèle de Shannon et Weaver qui représente l'acte de communication de façon imagée (fig. 1.1). À travers un canal, un émetteur transmet un message vers un récepteur. Ce message est soumis à des actions d'encodage (phase d'émission) puis décodage (phase de réception) qui sont soutenues par un code commun aux deux protagonistes. Dans le cas de la parole, cette représentation est incomplète dû à l'absence de l'expressivité. Ainsi, l'auteur modifie le modèle pour en proposer un nouveau en correspondance avec la communication vocale (fig. 1.2). Dans cette nouvelle représentation, l'expressivité apparaît

par l'ajout d'un deuxième couple d'encodage/décodage soutenu par le code para-linguistique. La communication vocale contient donc deux messages « qualitativement différents ». Le message primaire qui transforme, à l'aide du code linguistique, une idée en une suite de phonèmes et le message secondaire, soit l'expressivité, qui est généré par le corps humain lors de « l'acte de mise en sons des phonèmes ». Cette dissociation des deux messages met en lumière que l'expressivité s'inscrit sur le message primaire lors de la production des phonèmes. Iván Fónagy parle alors de style vocal, soit selon la définition courante : une « façon de parler, manière de s'exprimer ». Le message secondaire, présent dans la communication vocale, permet donc de prononcer une même phrase de différentes manières et crée ainsi des nuances du message primaire.

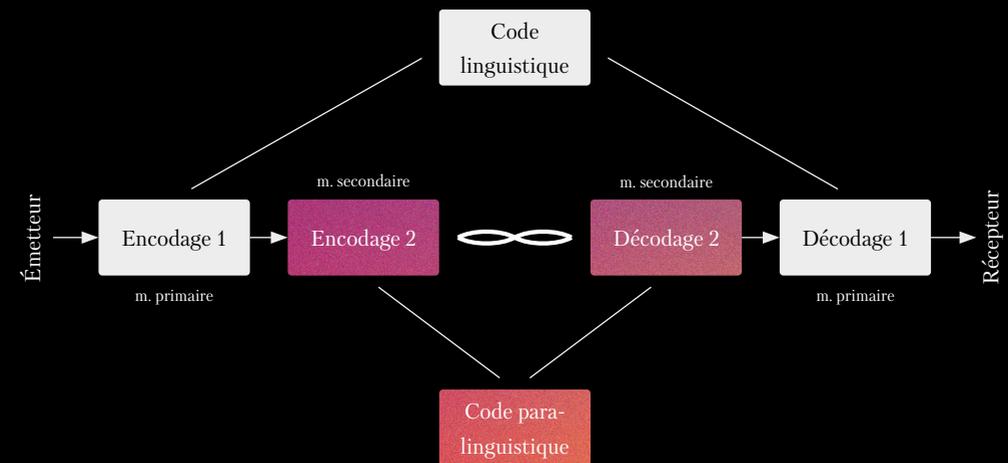
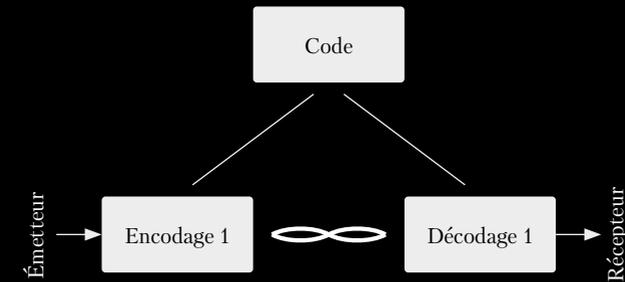


Figure 1.1. Illustration (en-haut)  
Représentation de la communication verbale.

Figure 1.2. Illustration (en-bas)  
Représentation de la communication vocale

Selon Iván Fónagy, le premier modèle (fig 1.1) est « trop général, trop schématique pour tenir compte du caractère spécifique des messages vocaux qui permettent de transmettre deux messages qualitativement différents ». L'auteur actualise le modèle (fig 1.2) de la manière qui suit : « il faudrait admettre, par conséquent, deux actes successifs d'encodage : un encodage linguistique qui

transforme un message global, une idée, en une séquence de phonèmes, et un deuxième codage - qui coïncide admirablement avec l'acte de mise en son des phonèmes - au cours duquel le message secondaire, gestuel est greffé sur le message primaire. Nous sommes amenés, néanmoins, à représenter graphiquement les deux actes d'encodage à l'aide de cases séparées ».



Figure 1.3. Illustration  
Le phonème /ɛ/ présent notamment dans les mots colère, mer, j'aimais

Avec l'ajout de ce second couple d'encodage/décodage (fig. 1.2), Iván Fónagy nous introduit deux facteurs qui participent à inscrire l'expressivité dans le message primaire. Le premier est lié aux phonèmes, le plus petit segment insécable de la phonologie<sup>1</sup>, au nombre de 36 dans la langue française (fig. 1.3). Par nature, un phonème dispose d'une marge dans la réalisation du son. Ainsi lors de la production sonore, il y a un choix qui est réalisé à partir de la forme pure du phonème. Ceci est notamment évoqué par Karl Bühler dont Iván Fónagy cite un extrait de *Lonomatopée* (1933, p.108) :

**« Chaque phonème laisse une certaine marge aux réalisations possibles, et c'est dans toutes ces marges que peut s'élaborer la peinture sonore. »**

Le second élément est lié au corps humain, responsable de « l'acte de mise en sons des phonèmes » par l'entremise des organes de la parole. Ainsi, c'est le corps qui s'exprime à travers les marges présentes dans chaque phonème émis. La somme de ces marges génère une nuance du message primaire. Toutefois, la question de l'expressivité n'étant pas entièrement résolue, attachons-nous désormais à comprendre quelles informations le corps confère au message.

<sup>1</sup> Définition du CNRTL (Centre national de ressources textuelles et lexicales)

## 1.2. La parole, reflet des émotions

Iván Fónagy en est convaincu, les émotions sont la principale source de l'expressivité. Celles-ci animent le corps humain qui va inscrire une nuance expressive au message primaire. Pour comprendre l'impact des émotions sur le corps, l'auteur nous renvoie à un passé lointain où les organes de la parole étaient avant tout destinés à la réalisation de fonctions biologiques autres, telles que la respiration. L'auteur invoque les travaux présentés dans l'ouvrage *The Expression of the Emotions in Man and Animals* (1872) de Charles Darwin où celui-ci replace les émotions dans le quotidien de l'homme primitif (fig. 1.4). Pour le scientifique anglais, les attitudes émotives comme la colère ou la haine permettaient de générer, à une époque ancestrale, des « activités dirigées vers des buts précis ». Ainsi, lors des combats ancestraux dont prenaient part les hommes primitifs, la colère et la haine envahissaient le corps afin d'apparaître plus menaçants. Ces émotions font alors naître une mimique particulière et instinctive sur le corps, le visage, mais aussi la

voix (annexe 1.1). Malgré un contexte différent, l'homme moderne n'échappe pas aux émotions et ses conséquences sur le corps. Chaque émotion, ayant sa somme de mouvements propres à elle (annexe 1.2), présentera des caractéristiques singulières sur la voix et la parole. Selon Iván Fónagy, ce jeu mimétique permettrait d'extérioriser une émotion et de la matérialiser.

Enfin, Iván Fónagy nous démontre que les émotions sont à l'origine de l'expressivité. Sous leurs influences, le corps humain génère une mimique particulière qui exploite la marge de chaque phonème. Le résultat produit est un ensemble d'écarts phonétiques par rapport à un état neutre qui inscrit l'expressivité sur le message.

La conversation étant une activité qui implique deux joueurs minimum, nous allons désormais nous attacher à comprendre comment le récepteur perçoit les émotions à travers la parole humaine.



Figure 1.4. Illustration  
Extrait du film *La guerre du feu*  
de Jean-Jacques Annaud (1981).  
L'extrait illustre la présence de la colère et  
de la haine afin d'apparaître plus menaçants

### 1.3. Entendre les émotions

Pour l'auteur de *La vive voix*, que ce soit sur le plan physique ou vocal, la perception des émotions se réalise par la comparaison inconsciente de l'état émotif à un état neutre. Ainsi l'être humain décompose et analyse chaque mouvement pour déceler un écart potentiel qui sera jugé comme une information en soi :

**« Ceci suppose une analyse préconsciente du son concret : sa décomposition. Le son est d'abord identifié à un des phonèmes vocaliques, puis la réalisation de ce phonème est comparée à l'articulation habituelle neutre, et l'écart articuloire licite une interprétation. L'auditeur reconstitue une deuxième composante latente, responsable de l'écart. Il considère la distorsion comme une expression. »**

L'auteur s'intéresse alors à la mélodicité de la parole humaine, fortement liée à l'expression des émotions. Dans la parole, la mélodie est définie comme « la courbe sonore résultant de la succession des intonations ou hauteurs des sons »<sup>1</sup> (fig. 1.4). Iván Fónagy nous rappelle que les successions de hauteur (fréquence en hertz) présentes dans la courbe mélodique sont perçus par l'être humain comme un mouvement. Ainsi, quand la fréquence augmente, le ton monte et l'être humain perçoit un mouvement allant vers le haut. Si la fréquence diminue, le mouvement suit.

Tout comme les mimiques du corps réalisées lors de l'émission, la courbe mélodique présente des mouvements singuliers pour chaque émotion et identifiables par le récepteur. Si nous reprenons le cas de la colère, le schéma mélodique de cette émotion a un profil angulaire (fig. 1.5) que l'auteur décrit comme tel :

**« Une ligne mélodique droite rigide, est interrompue à intervalles sensiblement égaux, dans les syllabes (fortement) accentuées, par des écarts brusques, d'une quarte, d'un quinte. Ce schéma montre une stabilité remarquable à l'intérieur de la même phrase, ou même au cours d'une séquence de phrase exprimant la même colère avec la même intensité. Les brusques sauts mélodiques atteignent à peu près la même hauteur, constituant un palier virtuel non moins rigide que la palier tonal de base. »**

S'il est désormais formel que les émotions sont retranscrites dans la parole humaine, attendons-nous à comprendre quelles autres informations y sont présentes.

<sup>1</sup> Définition du CNRTL

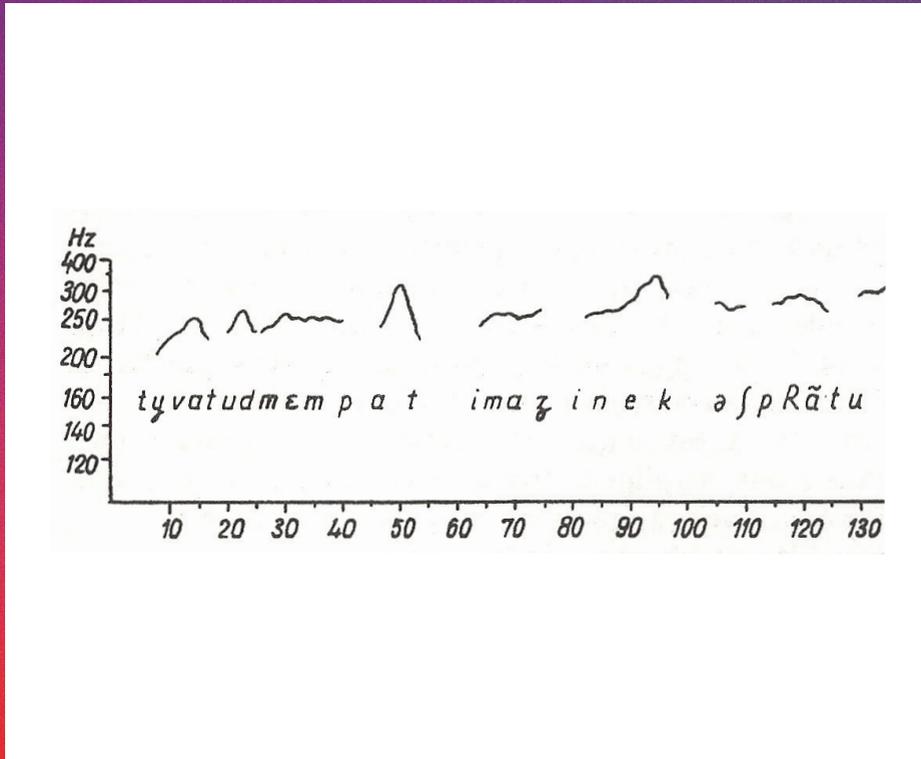


Figure 1.5. Graphique  
La courbe de fréquence fondamentale résultant d'une phrase dite avec colère

## 1.4. La voix, une empreinte vocale

Les travaux de recherches sur la parole humaine menés par Joana Révis et Jean Abitbol ont pour point de départ le caractère unique de chaque voix [Révis, 2013] [Abitbol, 2005]. Les deux auteurs sont formels : cette singularité est le fruit de la physiologie du système phonatoire, lui-même unique. D'un individu à un autre, la génétique opère sur le volume des poumons et du larynx, sur la longueur, le diamètre et la tension des cordes vocales et sur la forme de la bouche. La somme de toutes ces variations produisent un résultat propre à chacun, une voix avec une certaine fréquence fondamentale et telle une empreinte, la voix révèle notre identité comme en témoigne les auteurs :

**« La voix est une alchimie entre le corps et l'esprit. Elle révèle, dans nos silences, nos respirations et notre musicalité, les cicatrices de notre histoire de notre vie. Elle est donc l'image non seulement de notre personnalité, mais aussi de notre évolution. »**

Jean Abitbol

**« Parlez... et je vous dirai qui vous êtes. Notre voix dit beaucoup sur nous : âge et sexe, origines géographiques et socioculturelles, mais aussi physiologie, état de santé, ou émotions... »**

Joana Révis

Ainsi, c'est la hauteur de la voix (fréquence fondamentale) qui trahit le genre et l'âge. Des recherches ont permis de démontrer que la fréquence fondamentale d'une femme est près de 200 hertz (Hz), lorsque celle d'un homme avoisine les 125 Hz. Les hauteurs de voix sont donc généralement perçues comme plus graves pour les hommes et plus aiguës pour les femmes. À partir de la fréquence fondamentale, l'âge peut aussi être déterminé approximativement en attribuant à la voix une catégorie : enfant, adulte ou senior. De ce fait, les enfants ont une fréquence fondamentale proche de 300 Hz soit supérieure à celle d'une femme. En ce qui concerne la distinction entre la classe adulte et senior, celle-ci s'effectue à partir de 60 ans lorsque le processus naturel de dégénérescence du larynx, appelé presbyphonie, se produit. Ce phénomène physiologique a pour conséquence d'augmenter la fréquence fondamentale de la voix des hommes et de diminuer celle des femmes. Toutefois, l'âge ne se détermine pas uniquement par la hauteur de voix

comme nous l'indique Joana Révis :

**« On peut également concevoir que la voix reflète les caractéristiques générales que la société attribue aux individus en fonction de l'âge. Un sujet jeune devrait de ce fait posséder une voix tonique, spontanée, qui pourrait être caractérisée par un débit rapide, une large étendue vocale, un timbre clair, alors qu'un adulte plus mûr pourrait avoir une voix posée, plus réfléchie, avec un débit plus lent, une prosodie plus neutre. »**

Notre identité est aussi trahie par nos accents qui révèlent notre origine géographique et socioculturelle. Les linguistes parlent alors d'accent diatopique pour évoquer celui d'une région particulière tel que le célèbre accent provençal. L'accent diastratique est quant à lui lié à l'origine socioculturelle et se relève à l'aide de différentes caractéristiques de la parole comme le présente le tableau ci-dessous (figure 1.6) :

Figure 1.6. Tableau Résultats issus d'une recherche réalisée par Jones et McMillan en 1973 concernant l'accent diastratique lié à l'origine socioculturelle.

“Bas” niveau socioculturel	“Haut” niveau socioculturel
Articulation relâchée	Diction soignée
Débit rapide	Débit lent
Prosodie étendue	Prosodie modérée
Accent marqué	Accent “standard” à “pointu”
Vocabulaire “dévoté”, argot	Vocabulaire académique, soutenu

## En conclusion,

les trois auteurs que nous avons étudiés dans ce chapitre sont explicites sur le fait que la parole humaine, par nature, porte des informations difficilement accessibles à l'écrit. Iván Fónagy évoque la parole sous les caractéristiques de la vivacité et de l'expressivité. Ses travaux de recherches décompose l'acte de parler pour atteindre la source de ces caractéristiques qui n'est autre que les émotions. Celles-ci s'inscrivent sur chaque phonème, chaque mot et seront perçues par le récepteur à travers la mélodie. Quant à Joana Révis et Jean Abitbol, leurs travaux de recherches prennent leur source dans le caractère unique de chaque voix. Telle une empreinte, la voix révèle notre genre, notre âge et nos origines.

Si notre recherche s'est principalement orientée sur les informations présentes dans les aspects

vocaux et phonétique, la dimension linguistique n'est pour autant pas en reste. Lorsque l'être humain s'exprime, il sélectionne son vocabulaire, ses expressions et la forme de la syntaxe, un tout qui donne de nouvelles informations. Ceci met en perspective le vaste sujet qu'est la parole humaine.

En ce qui concerne la problématique de ce mémoire, si l'empathie est «la capacité de s'identifier à autrui, d'éprouver ce qu'il éprouve»<sup>1</sup>, une des réponses est sans nul doute la présence des émotions dans la parole humaine. Évoqué par Iván Fónagy, l'être humain, par nature, perçoit et identifie au travers de la voix les émotions de ses pairs.

---

1 Définition du CNRTL

# La parole dans les technologies

Comme les traits dans les camées  
J'ai voulu que les voix aimées  
Soient un bien, qu'on garde à jamais,  
Et puissent répéter le rêve  
Musical de l'heure trop brève ;  
Le temps veut fuir, je le soumets.

Charles Cros

Le collier de griffes, *Inscription*, 1908

Ce deuxième chapitre portera un regard sur la place de la parole humaine dans les technologies passées et actuelles. Les vers extraits du poème *Inscription* (ci-contre) de Charles Cros, inventeur du paléophone<sup>1</sup>, illustre sa volonté de mener la parole au-delà des capacités spatio-temporelles permises par le corps humain. Au cours du XIXe siècle, cette volonté d'explorer les possibilités de la parole humaine n'était pas unique à l'inventeur français.

Plusieurs de ses pairs ont conduit des recherches aboutissant sur la découverte de machines qualifiées de parlantes.

Ainsi, nous débiterons ce chapitre par une présentation des travaux menés par ces inventeurs du XIXe siècles pour remonter jusqu'à notre période actuelle. L'objectif est de mettre en lumière les zones à explorer en ce qui concerne la parole humaine dans le numérique.

---

<sup>1</sup> Le paléophone est un appareil de reproduction des sons, cousin du phonographe de Thomas Edison.

## 2.1. Les machines parlantes

C'est au cours du XIXe siècle, ère des grandes inventions, que naissent les machines parlantes. Leurs découvertes sont l'oeuvre de quelques inventeurs qui souhaitaient transmettre la parole au-delà des capacités physiques permises par la voix humaine. En 1876, Alexander Graham Bell offrit à l'être humain la possibilité de communiquer avec un pair instantanément à travers le monde en concevant le téléphone. En 1896, Guglielmo Marconi donna à l'être humain la possibilité de transmettre simultanément sa parole dans le monde entier en combinant trois précédentes découvertes (l'antenne, les ondes Hertiennes et la radioconduction) et dont le résultat fut la radio. Deux inventions qui transgressent les limites spatiales de la parole humaine.

Le phonographe, quant à lui, redéfinit les capacités spatio-temporelles de la communication vocale. Par le génie de Thomas Edison, l'être humain pu pour la première fois s'enregistrer et écouter *a posteriori* ces « voix du passé ». Dans un article du *North American Review* publié quelques mois après le dépôt du brevet [Edison, 1878], Thomas Edison y décrit les multiples usages du phonographe (annexe 2.1) tels qu'apprendre à jouer des morceaux de musique, écouter *a posteriori* la leçon d'un cours ou encore enregistrer les dernières paroles d'un membre de la famille. Toutefois, l'inventeur atteste que le phonographe a, en premier lieu, été conçu pour la lettre sonore et projette l'idée d'êtres humains communiquant par enregistrements vocaux. Thomas Edison imagine aussi la possibilité de lire des livres audio avec le phonographe. La littérature serait ainsi accessible aux non-voyants, mais aussi aux personnes qui souhaiteraient réaliser simultanément

plusieurs activités. À la lecture des usages du phonographe, il est étonnant de constater leurs similarités avec des usages qui nous sont contemporains comme le concept de *multitasking* démocratisé avec la technologie numérique.

**“C'est une impression vraiment saisissante que celle que l'on ressent la première fois que l'on entend parler le phonographe. Cette voix grêle, ce timbre légèrement métallique, ce ton nasillard, ce chant contourné de la phrase produit par un mouvement irrégulier du cylindre, - tout cela prête une vie étrange à ces organes mécanique. Il semble qu'un esprit moqueur s'amuse à répéter en caricature tout ce qu'on vient de dire.”**

Antoine Breguet (1878)

Au regard de ces objets, leurs inventeurs ont intégré des attributs qui font, de nos jours, référence à la communication vocale. Ce qui lors des découvertes pouvait paraître surprenant nous est désormais acquis : nous parlons dans le micro et nous écoutons dans le haut-parleur. Deux actions inhérentes à la communication vocale. Ainsi, le téléphone intégrant ces deux attributs nous signifie la possibilité de parler et d'écouter (fig 2.1). Quant à la radio, ne possédant que le haut-

parleur, celle-ci nous indique notre statut de récepteur (fig. 2.2).

Dans les lignes qui suivent, nous projeterons notre regard sur un des objets emblématiques du XXe siècle : l'ordinateur, avec le constat que celui-ci ne possède à la naissance aucun de ces deux attributs. De ce fait et par lui-même, l'ordinateur signifie son incapacité à gérer les usages de la communication vocale (fig 2.3).

Figure 2.1. Illustration Photographie prise lors du premier appel téléphonique réalisé par Alexander Graham Bell.





Figure 2.2. Illustration  
La radio Braun RT 20 conçu par Dieter Rams (1961).  
La radio est notamment composée d'un haut-parler signifiant notre statut de récepteur.



Figure 2.3. Illustration  
L'ordinateur Apple II conçu par Steve Wozniak (1977)  
Les premiers ordinateurs n'étaient pas équipés de micro et de haut-parleurs  
signifiant leur incapacité à gérer les communications vocales.

## 2.2. L'ère informatique

C'est autour des années 1939 à 1945 que naissent les premiers ordinateurs et pour cause, ce sont les besoins militaires qui exigent de grandes avancées scientifiques dans le domaine des machines programmables. L'un des fait de guerre les plus connus est le décodage du système de chiffrement allemand Enigma. C'est dans le centre secret de la *Government Code and Cypher School* (GC&CS) à Betchley Park (fig. 2.4) qu'une équipe de scientifiques britanniques, dont le célèbre mathématicien Alan Turing, tentait de déchiffrer le système allemand. Pour ce faire, ils conçoivent d'importants calculateurs, ancêtres de nos ordinateurs actuels, capables de tester en série des clés de chiffrement. Ces machines programmables permettaient en une journée de réaliser le travail de milliers de personnes. Ainsi, c'est par la fonction de calculateur que naissent les premiers ordinateurs et il en sera de même pendant de nombreuses années. Aujourd'hui encore, les ordinateurs nous permettent de réaliser quantité d'opérations et de tâches, augmentant ainsi les capacités cognitives de l'être humain. L'exploitation de celui-ci comme outil de communication ne s'effectuera réellement qu'avec la démocratisation d'internet. Les ordinateurs s'apparentent donc à des machines à écrire où tout repose sur une interface graphique (fig. 2.3). La parole humaine y est absente.

Toutefois, quelques travaux scientifiques ouvrent la possibilité d'interagir vocalement avec la machine. Ainsi, en février 1979, alors que les ordinateurs personnels se démocratisent, un article de *Science* pose la question de l'utilisation de la voix comme interface homme-machine [Robinson,

1979]. Selon la revue américaine, une telle interface porte un avantage majeur : la facilité et la flexibilité de la conversation, plus simple à utiliser que la rigidité du clavier et du langage informatique. De plus, l'efficacité de la communication orale permet d'aller plus vite et offre la possibilité de réaliser plusieurs tâches en même temps. L'année suivante, Richard A. Bolt, membre du Architecture Machine Group du MIT, publie ses recherches (*Put that there*) sur la combinaison de la voix et de la gestuelle pour exécuter des tâches visibles sur une interface graphique [Bolt, 1980]. Dans une courte vidéo<sup>1</sup>, le professeur interagit avec son ordinateur et lui ordonne par la parole et à l'aide d'un pointeur de déplacer et de modifier des formes géométriques présentes à l'écran. Ces premiers travaux sur les interfaces vocales peuvent s'apparenter aux prémisses de nos assistants vocaux qui n'auraient peut-être jamais vu le jour sans le complément d'une question posé par Alan Turing :

### «Can machines think?»

Pour répondre à cette question, le mathématicien anglais conçoit un exercice baptisé le jeu de l'imitation [Turing, 1950] Les règles, simples, proposeraient à un être humain d'échanger avec deux autres joueurs pour tenter de deviner lequel est une machine. Selon le mathématicien, l'acte de conversation nous permet de nous distinguer de la machine parce qu'on y décèle l'intelligence humaine. Ainsi, naît l'intelligence artificielle, autre technologie apparentée aux assistants vocaux.

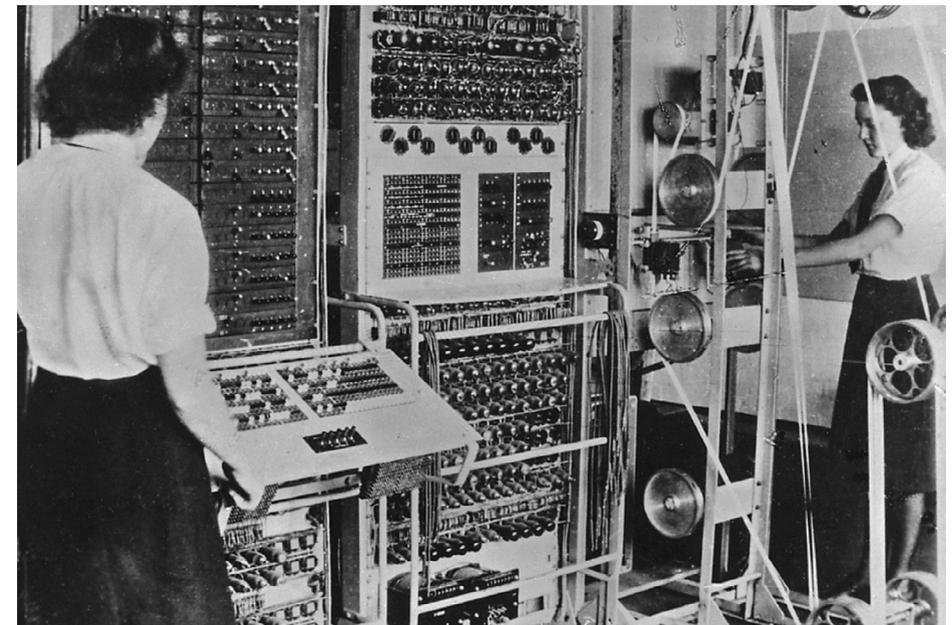
Toutefois, sans lien de parenté avec les machines parlantes, l'ordinateur n'a jamais développé de terrain favorable pour permettre l'émergence de la parole humaine dans la technologie numérique. La donne change en 2007<sup>2</sup> avec l'arrivée d'un objet ambigu : le smartphone. Mi-ordinateur mi-téléphone, celui-ci conserve les attributs que sont

le micro et le haut-parleur de son ancêtre filaire. Ces deux composants permettent au téléphone intelligent de signifier sa capacité à gérer les usages liés à la communication vocale. Ainsi, la combinaison de ces deux objets a-t-il généré un retour des usages liés à la parole humaine ?

<sup>1</sup> Put-That-There, 1982. MIT Media Lab. Youtube

<sup>2</sup> La première génération d'iPhone est présentée par Steve Jobs le 9 janvier 2007

Figure 2.4. Illustration  
Photo du Colossus prise en 1943 à Betchley Park, en Angleterre.  
Cet ordinateur a été conçu au cours de la seconde guerre mondiale afin de décrypter les codes allemands



## 2.3. La parole dans le numérique

C'est en 2011 que se concrétise pour le grand public un fantasme vieux de quelques décennies : interagir vocalement avec les machines (fig. 2.5). C'est Apple qui ouvre la voie avec l'intégration de Siri dans l'iPhone<sup>1</sup>. Dix ans plus tard, c'est la quasi-totalité des GAFAM<sup>2</sup> qui nous proposent de nous libérer de nos interfaces graphiques pour accéder aux informations disponibles sur internet (actualités, météo, etc.) et pour communiquer avec nos pairs (lecture et envoi de messages textes, etc.)<sup>3</sup>.

Dans le cadre de leurs recherches, Ewa Luger et Abigail Sellen, deux chercheurs de Microsoft, ont étudié entre autres les motivations liées à l'utilisation des assistants vocaux [Luger et al, 2016]. Selon les deux chercheurs, une des raisons majeures est le gain de temps. Les utilisateurs ont le sentiment de réaliser moins d'étapes lorsqu'ils effectuent certaines tâches, comme l'ajout d'un rendez-vous ou la programmation d'un réveil. Ce sentiment d'efficacité naît aussi de la possibilité d'effectuer plusieurs activités en simultané. Néanmoins, quelques ombres au tableau sont aussi révélées à l'usage telles que la nécessité de simplifier ses phrases et d'énoncer les bons mots-clés pour que la requête soit comprise et réalisée par l'assistant. Cette performance fluctuante des assistants vocaux fait douter certains individus de l'étude qui de fait ne réalisent pas certaines actions critiques comme dicter un message. Malgré cela, la commande vocale se démocratise et les GAFAM nous proposent désormais d'équiper nos

pièces de vie d'enceintes connectées, marquant le retour d'objets dédiés à la parole. Cette instrumentalisation de la parole semble donc conditionner l'être humain à interagir vocalement avec la machine, mais pas seulement.

C'est ce même gain de temps qui anime certaines personnes à envoyer des messages vocaux et non textuelles à ses pairs. Ce nouveau mode de communication est baptisée le *voice messaging* et émerge de certaines régions du monde, notamment la Chine, ainsi que de certaines générations. C'est par l'intermédiaire d'une application de messagerie instantanée (Whatsapp, Messages, Messenger, ...) qu'un correspondant peut envoyer un message vocal à un ou plusieurs destinataires. Pour ce faire, il lui suffit de tenir appuyer le bouton d'enregistrement, souvent représenté par un micro, et de parler (fig. 2.6). Si le gain de temps est l'une des portes d'entrée du *voice messaging*, plusieurs utilisateurs mentionnent d'autres avantages. Un article du Figaro<sup>4</sup> évoque sur le sujet « l'envie de transmettre des émotions ou des informations de façon plus authentique », là où les messages textes, même accompagnés d'émoticônes, conservent une certaine ambiguïté, notamment pour le ton du message. Le professeur Benjamin Cowan, dont les recherches questionnent les interfaces homme-machine, relate dans un article du journal britannique *The Guardian*<sup>5</sup> un sentiment de connexion plus fort avec ce nouveau mode de communication :

<sup>4</sup> Un article rédigé le 2 janvier 2019 par Harold Grand pour le Figaro : Les messages vocaux s'imposent dans le smartphone des adolescents

<sup>5</sup> Un article rédigé le 2 décembre 2019 par Chris Stokel-Walker pour *The Guardian* : *Voice messaging – conversational gain or pain?*

<sup>1</sup> Siri est présenté le 4 octobre 2011 lors de la Keynote d'Apple.

<sup>2</sup> Google, Apple, Facebook, Amazon et Microsoft

<sup>3</sup> Page web de l'assistant vocal Siri (Apple)

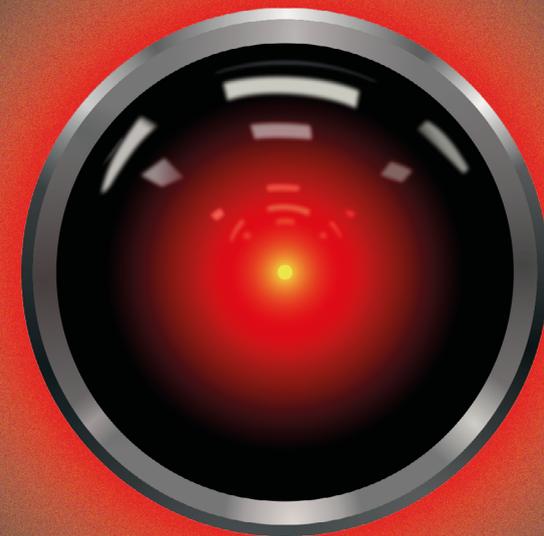


Figure 2.5. Illustration

Extrait du film 2001, *l'Odysée de l'espace* de Stanley Kubrick (1968) mettant notamment en scène HAL 9000, intelligence artificielle dotée de la parole.

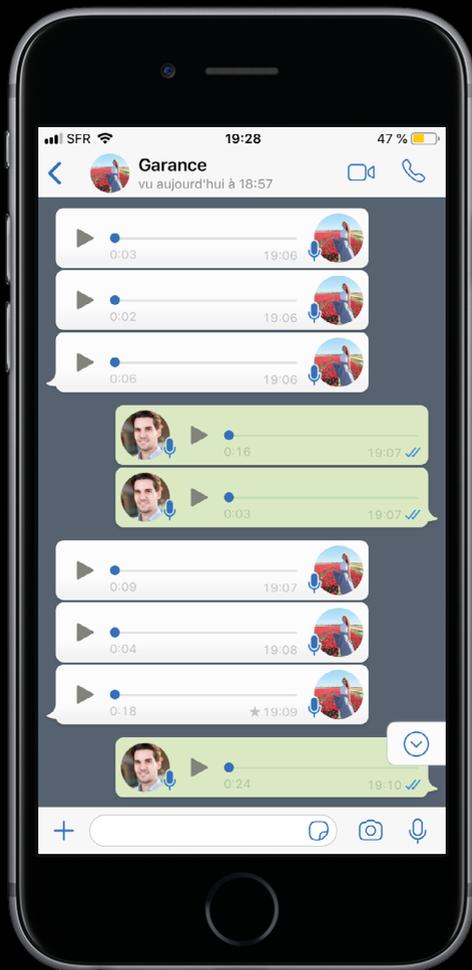


Figure 2.6. Illustration

Une conversation dans l'application WhatsApp.

Les messages vocaux sont représentés par des lecteurs audio distincts, intégrant une ligne temporelle. Cette représentation ne permet pas d'avoir accès visuellement au contenu du message.

**« It gives an honest impression of what you're wanting to say at the time. You can be more emotive and more spontaneous with it than text-based communication. [...] Voice communication gives you a really rich sense of emotion and a higher sense of connection. »**

**Benjamin Cowan**

Dans son article *Voice messaging – conversational gain or pain?*, The Guardian<sup>1</sup> questionne les avantages et inconvénients de cet outil de communication. Pour le professeur Benjamin Cowan, il semble indéniable que le *Voice messaging* simplifie l'action de converser pour l'émetteur, mais au détriment des destinataires. À la réception du message vocal, les destinataires ne peuvent, comme pour un message écrit, «lire» en

1 Un article rédigé le 2 décembre 2019 par Chris Stokel-Walker pour *The Guardian* : *Voice messaging – conversational gain or pain?*

diagonale le contenu et en comprendre la teneur. Est-ce une bonne nouvelle ? Une urgence à traiter ? Impossible à dire avec la forme actuelle des messages vocaux représentés par un lecteur audio accompagné d'une ligne de temps (fig. 2.6). Pour les récepteurs, les contraintes ne s'arrêtent pas là, ces derniers devront pour écouter l'enregistrement audio dans de bonnes conditions ou en toute confidentialité se placer dans un lieu approprié ou s'équiper d'un casque audio. D'autre part, une première écoute peut ne pas être suffisante pour obtenir l'ensemble des informations, obligeant les récepteurs à écouter plusieurs fois intégralement un même message. Les interfaces graphiques des messageries instantanées ne proposent pour le moment aucune information en support de l'écoute.

Pour autant, ces contraintes n'empêchent pas une forte croissance de l'utilisation des messages vocaux<sup>2</sup>. En Chine, plus de 90% des utilisateurs de WeChat exploitent le *voice messaging* atteignant 6,1 milliard de messages vocaux par an. L'application Viber enregistre une hausse de 50% en 2018. Quant à Messenger, Facebook affirme que les messages vocaux sont désormais le deuxième type de fichier le plus envoyé sur sa plateforme, derrière les images<sup>2</sup>. Ainsi, une majorité des messageries instantanées enregistrent des records concernant ce nouveau mode de communication. L'engouement est tel que ces entreprises évoquent dans la stratégie produit de nouvelles fonctionnalités comme la retranscription vocale. Dans un avenir proche, les utilisateurs pourront donc écouter et/ou lire un message réceptionné.

2 Un article rédigé le 2 janvier 2019 par Harold Grand pour le *Figaro* : Les messages vocaux s'imposent dans le smartphone des adolescents

## En conclusion,

l'arrivée du smartphone accompagnée de nouvelles technologies (intelligence artificielle, reconnaissance vocale, etc.) semble signer le grand retour des usages et objets liés à la parole humaine (assistants vocaux, voice messaging, podcast, VoIP<sup>1</sup>). En ce qui concerne l'ordinateur, malgré ses origines de calculateur, celui-ci n'est pas en reste et intègre désormais les assistants vocaux. Connectée en tout temps, la machine est exploitée pour effectuer des appels téléphoniques, de la visioconférence ou communiquer via les réseaux sociaux. Ainsi, ce regard porté sur la place de la parole humaine dans les technologies passées et

actuelles semble mettre en lumière un moment favorisant le développement de nouveaux usages liés à la parole.

En réponse à la problématique de ce mémoire, nous apprenons que l'une des raisons de l'usage du *voice messaging*, au détriment du message texte, est « l'envie de transmettre des émotions »<sup>2</sup>, le besoin d'une connexion plus forte avec ses pairs<sup>3</sup>. Ainsi, certains utilisateurs du *voice messaging* semblent s'accorder sur le fait que la parole humaine et enregistrée favorise des interactions humaines plus empathiques dans le numérique.

---

1 Acronyme de Voix sur IP

2 Un article rédigé le 2 janvier 2019 par Harold Grand pour le Figaro : Les messages vocaux s'imposent dans le smartphone des adolescents

3 Un article rédigé le 2 décembre 2019 par Chris Stokel-Walker pour *The Guardian* : *Voice messaging – conversational gain or pain?*

« It's through mistakes  
that you actually can grow.  
You have to get bad  
in order to get good. »

Paula Scher

Chapitre 03

---

# Comment pourrions-nous ?

Les deux premiers chapitres de ce mémoire ont mis en évidence la richesse de la parole humaine et l'engouement récent des entreprises technologiques pour ce mode de communication. Il est donc nécessaire de saisir ce *momentum* favorisant le développement de nouveaux usages liés à la parole pour proposer des objets qui ne vont pas dans le sens de l'intermédiation par la technologie numérique. Cette étude nous convainc donc de poursuivre la réflexion sous une autre forme : faire projet. L'idée

de ce troisième chapitre est de sortir de l'abstraction par la conception d'un cas réel, présenté dans les pages qui suivent. Pour autant, dans le cadre du mémoire la réalisation complète d'un projet n'est pas une fin, c'est le processus que nous observons. L'objectif de la démarche est d'explorer les ouvertures qu'offre la problématique, d'imaginer de nouveaux usages afin de consolider le mémoire et de l'ancrer dans une perspective de projet innovant.

### 3.1. Recherche d'inspirations

Exprimé en introduction de ce mémoire, le sentiment d'intermédiation de la technologie numérique dans les relations humaines me conduit à débiter la réponse par une phase de recherche d'inspirations. Pour ce faire, je me mets en quête d'images, de lieux et de situations où les échanges entre êtres humains ont une place fondamentale, afin d'identifier les composants qui favorisent la conversation orale. De cette première phase, je choisis volontairement d'exclure la période de révolution numérique qui tend à placer la technologie comme intermédiaire des relations humaines.

Ma recherche me fait très rapidement remonter jusqu'à la Grèce antique dont les images d'orateurs, de philosophes et de la démocratie naissante retiennent mon attention (fig. 3.1). Naît alors dans mon imaginaire un fantasme d'une époque passée où la parole avait une place telle dans la vie des citoyens grecs, qu'ils érigeaient des lieux pour soutenir celle-ci, comme l'agora. Imaginant l'agora comme le lieu où se réunit l'assemblée de citoyens grecs, j'effectue une brève recherche concernant cet édifice emblématique de la Grèce antique. Ainsi, le CNRTL définit l'agora comme un « lieu où à l'origine, se réunit l'assemblée des citoyens, puis qui devient la place principale de la cité : centre politique, économique, religieux autour duquel se range une partie des édifices publics »<sup>1</sup> et dont l'étymologie est issue du mot grec ancien ἀγορά<sup>2</sup> que l'on traduit par :

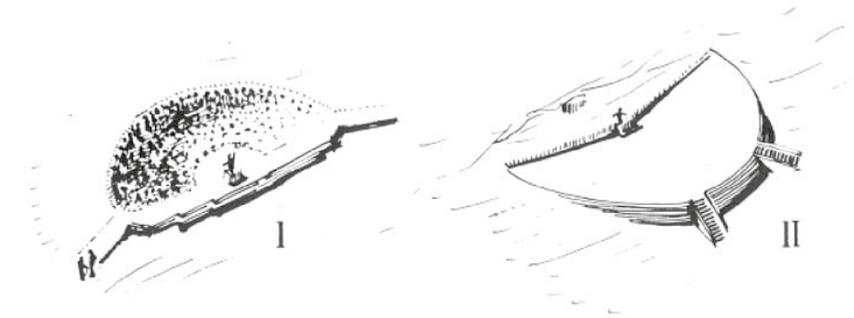
« lieu où  
l'on se réunit »

Ce qui m'interpelle dans la racine du mot et sa définition, c'est que l'agora est finalement bien plus qu'une simple assemblée de citoyens imaginée. C'est un lieu unique, centralisant diverses activités, ouvert à tous, permettant ainsi de se retrouver et d'échanger quotidiennement. Dans une certaine mesure, ces caractéristiques ne sont pas sans me rappeler le pendant romain de l'agora : le forum. Ce même forum qui dans le milieu informatique est désormais un espace de discussion ouvert à tous et permettant d'échanger sur des sujets divers et variés. La poursuite de recherches sur la Grèce antique me conduit finalement à retrouver le nom de l'assemblée de citoyens grecs : l'ecclésia et dont le siège résidait non loin de l'agora sur la colline Pnyx. Sur le site officiel des fouilles archéologiques de l'agora d'Athènes<sup>2</sup>, je découvre une planche de trois dessins présentant l'évolution architecturale de l'auditorium de l'ecclésia (fig. 3.2).



Figure 3.1. Illustration  
Le Tableau *Oraison funèbre de Périclès* de Philipp von Foltz (1852)  
La prise de parole de Périclès devant l'auditoire composé de citoyens athéniens à l'ecclésia.

Figure 3.2. Illustration  
Évolutions architecturales du siège de l'ecclésia situé sur la colline Pnyx  
Dessins de John Travlos, architecte en charge des fouilles de l'agora d'Athènes  
Athenian Agora Excavations - American School of Classical Studies at Athens



<sup>1</sup> Définition du CNRTL

<sup>2</sup> Site internet de Athenian Agora Excavations : agathe.gr

## 3.2. Représentation spatiale

De cette recherche germent mes premières idées reposant sur des éléments que je détaille dans les lignes qui suivent.

À la vue de la scène peinte par Philipp Von Foltz (fig. 3.1), je retiens la mise en lumière de l'orateur, ici Pericles, qui est isolé de l'auditoire par des artifices artistiques et sa position, droit et seul sur l'estrade. Démontrée par les fouilles<sup>1</sup>, la réalité était toute autre : l'orateur n'était pas entouré par son auditoire, mais de par la position centrale de l'estrade il faisait face à lui.

Des dessins de John Travlos (fig. 3.2), je retiens la forme en demi-cercle de l'auditorium qui définit les frontières du lieu ainsi que sa capacité d'accueil et qui n'est pas sans rappeler la forme de l'onde. Dans le cas de l'ecclésia, les frontières limitant l'accès au lieu ne sont pas uniquement physiques, mais aussi sociales. Ainsi, seuls les

hommes de citoyenneté grecque pouvaient entrer dans l'auditorium de l'ecclésia.

Je retiens aussi l'exploitation par les grecs des techniques architecturales dans le but de soutenir la parole et de la porter à l'ensemble de l'auditoire, comme l'évoque Vitruve<sup>2</sup>, célèbre architecte romain :

**« ils ont su faire en sorte que les voix provenant de la scène parviennent plus distinctement et plus agréablement aux oreilles des spectateurs [...] en concevant des théâtres selon les lois de la science de l'harmonie, les anciens ont accrue la puissance sonore de la voix »**

Ainsi, l'architecture se met au service de l'humain et de ses idées. Pour autant, l'orateur, mis sur un piédestal, semble isolé de l'auditoire. Au cours de ma réflexion, je réalise une représentation spatiale des différentes composantes de l'ecclésia (fig. 3.3) qui me conduit vers cette question : l'architecture de l'ecclésia permettait-elle vraiment d'inciter les échanges, d'ouvrir le dialogue ou inversement de favoriser les discours, les monologues et les idées d'un seul homme ?

Figure 3.3 - Illustration (à gauche)

Représentation spatiale de l'ecclésia et de ses composantes :

[1] L'auditorium qui exploite les techniques architecturales pour soutenir la parole et dont l'enceinte du lieu définit sa capacité.

[2] L'auditoire qui est défini par les droits d'accès liés aux règles sociales en vigueur.

[3] La scène qui surélève l'orateur et le détache de l'auditoire.

[4] L'orateur qui délivre son discours composé de multiples idées.

1 Site internet de Athenian Agora Excavations : [agathe.gr](http://agathe.gr)

2 Un article rédigé le 2 avril 2007 par Nico E. Declercq pour Futura Sciences : D'où provient l'acoustique extraordinaire du théâtre antique grec ?

De cette question, je réalise une deuxième représentation spatiale (fig. 3.4). Exit la scène, la mise en valeur d'un seul homme et le terme d'auditoire. Cette version met tous les protagonistes sur un pied d'égalité. Les frontières sont toujours présentes, mais plus anecdotiques. Les protagonistes semblent pouvoir y entrer et sortir à leur guise.

Par ailleurs, l'idée de lieu géographiquement localisé ne semble pas être une caractéristique importante. Le lieu permet l'échange mais c'est avant tout les interlocuteurs qui représentent la discussion. Ainsi, chacun peut partager librement ses idées, le tout dans le respect des conventions implicites de la conversation.

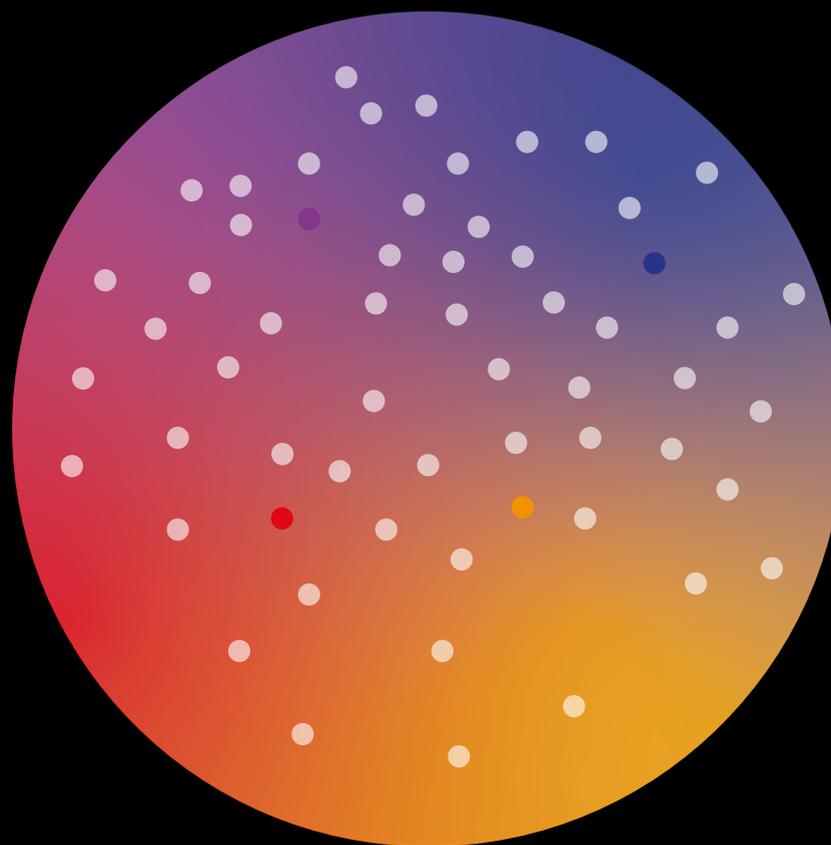
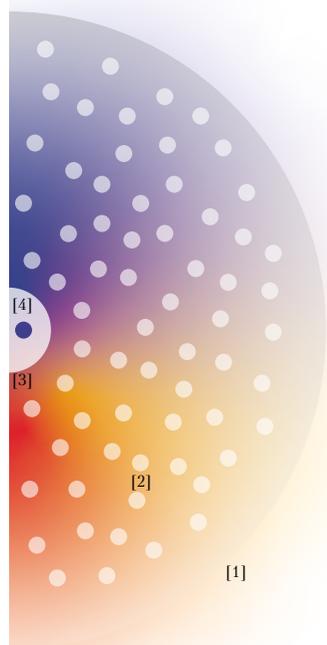


Figure 3.4. Illustration  
Représentation spatiale d'une conversation composée de plusieurs protagonistes qui partagent les uns après les autres leurs idées.

### 3.3. Première version

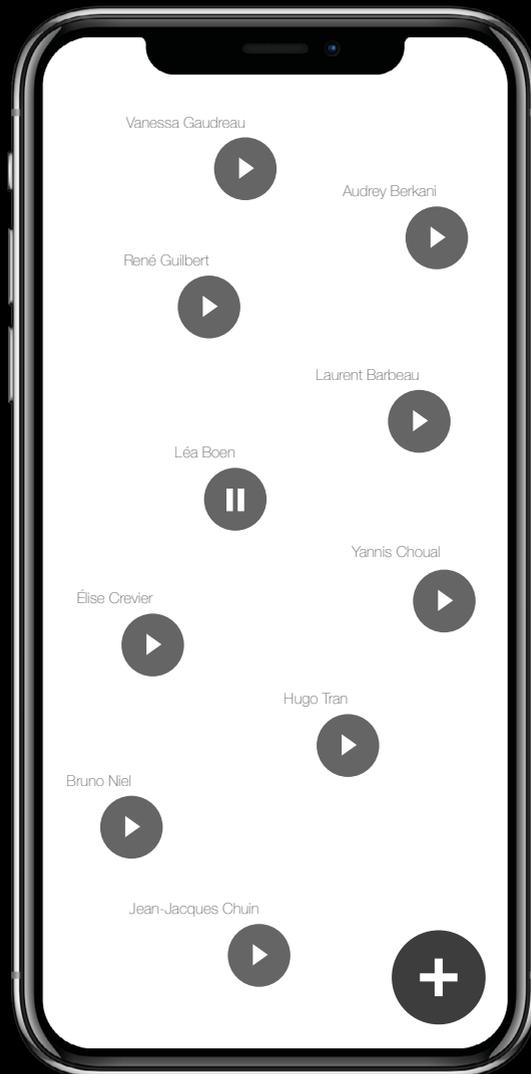


Figure 3.5. Illustration  
Esquisse d'Agoraphone avec une vue sur  
l'ensemble des enregistrements présents dans l'application.

Ces édifices emblématiques de la Grèce antique me font réaliser la nécessité de construire un lieu virtuel, ouvert à tous et où la technologie est exploitée uniquement pour être au service de la parole humaine et des échanges. Afin de me projeter dans cette idée, je retiens le nom d'Agora, que je juge plus inclusif que l'éclésiastique qui est restreint aux activités politiques. Je le complète du suffixe -phone pour évoquer la notion de son. Le projet qui va suivre sera donc présenté sous le nom d'Agoraphone. Partant du concept de lieu, je conserve l'idée de représenter spatialement les personnes et leurs contributions vocales (fig. 3.4). J'esquisse alors quelques dessins d'un lieu virtuel où tout un chacun pourrait s'exprimer oralement en publiant un enregistrement audio. D'autre part, au vu de l'analyse de la parole dans les technologies actuelles (chapitre 2), concevoir en premier lieu des interfaces mobiles semble une évidence. Le smartphone intégrant un micro et un haut-parleur induit des usages liés à la communication vocale. L'ordinateur n'est toutefois pas à négliger et peut-être une bonne alternative dans certains environnements comme le bureau. Dans cette première ébauche présentée par l'es-

quisse ci-contre (fig. 3.5), les enregistrements audio sont représentés par la forme d'un lecteur audio circulaire composé uniquement d'une touche lecture/pause et annoté du nom de l'auteur. Lorsqu'un utilisateur souhaite ajouter un enregistrement audio, celui-ci doit cliquer sur le bouton « plus » présent en bas à droite. À ce stade, les frontières de ce lieu virtuel sont délimitées uniquement par le nom de domaine du site internet. Ainsi, tous les enregistrements audio publiés sur la plateforme sont disponibles à l'ensemble des utilisateurs d'Agoraphone. Toutefois, si la situation nécessite de la confidentialité, la technologie numérique peut nous permettre de restreindre les frontières à une sphère privée, telles que la famille ou une équipe de travail.

Dans cette première version (fig. 3.5), la réponse à la problématique est incomplète car l'Agoraphone ne permet pas à l'auditoire de répondre aisément et directement à un orateur, mais seulement de l'écouter. La conséquence serait un sentiment de rupture toujours présent et une frustration liée à l'impossibilité de répondre directement à un interlocuteur.

### 3.4. Deuxième version

Donner la parole à l'auditoire ou répondre à une personne n'est pas anodin et repose dans la vie réelle sur des mécanismes conversationnels implicites que nous étudierons plus tard (section 3.7). Pour cette raison, je décide de mettre de côté la conception du parcours permettant de répondre à un message audio et de me concentrer sur la question de comment représenter la conversation .

Dans un premier temps, afin de susciter un échange entre différents protagonistes, je décide de rendre visible les interlocuteurs et de créer un premier lien visuel entre le locuteur et les protagonistes qui l'écoutent (fig. 3.6). Tout comme, le tableau de Philipp von Foltz (fig. 3.1), je mets graphiquement en lumière la personne qui parle à l'instant T avec l'ajout d'un halo de couleur symbolisant l'onde sonore de la parole (fig. 3.6). Je poursuis la conception avec l'idée réduite

qu'une conversation n'est autre que des protagonistes qui prennent la parole les uns après les autres. Ainsi, s'ajoute la notion de temps et d'interventions qui se suivent linéairement. Je transforme alors les lecteurs audio de la version précédente (fig. 3.5) en un objet unique et à part entière qui intègre une ligne de temps. De cette manière, il est possible d'écouter la conversation au complet ou partiellement et de comprendre la suite logique des idées et des propos avant d'ajouter sa réponse.

Cette première réponse à la question de représentation de la conversation m'amène à considérer la pertinence de représenter celle-ci spatialement. En cause, des informations graphiques qui évoluent peu, quand la conversation est supposée être riche d'idées. Cette réflexion me conduit à une autre question : n'est-il pas naturel de se représenter une conversation temporellement ?

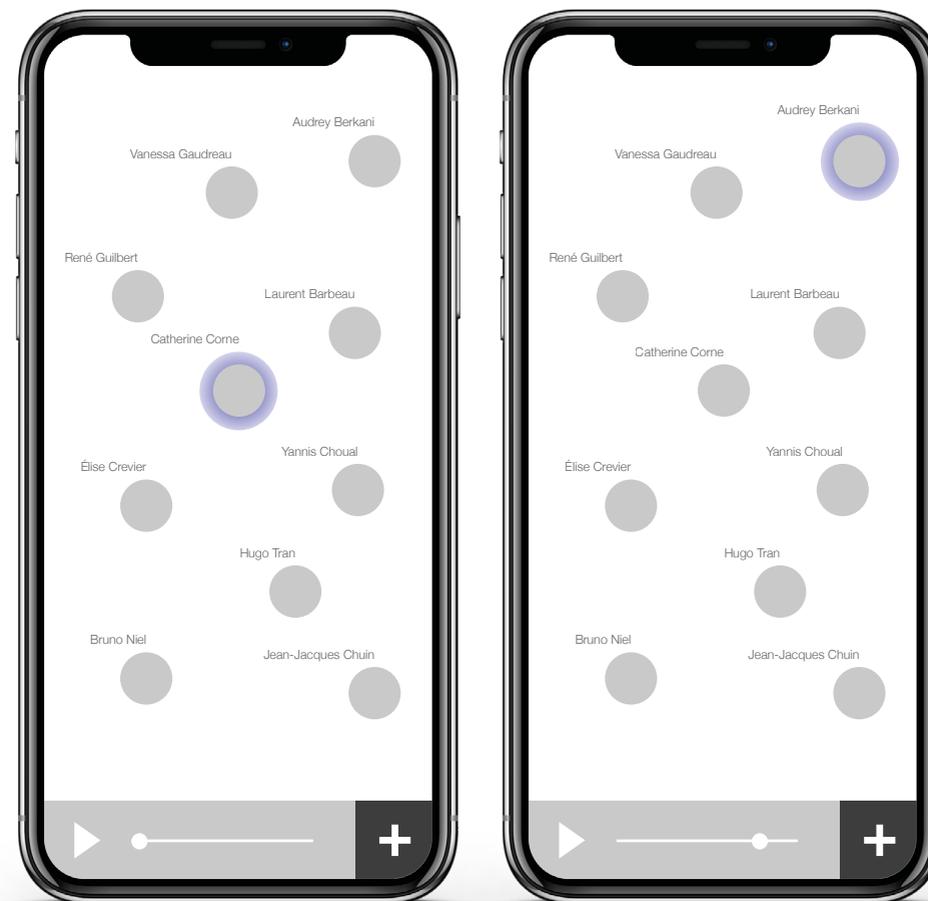


Figure 3.6. Illustration.  
Esquisse d'Agoraphone avec l'orateur et ses interlocuteurs.  
À gauche, le propriétaire de la conversation qui initie la conversation.  
À droite, un protagoniste de l'auditoire qui intervient à son tour, plus tard dans la conversation.

## 3.5. Représentation temporelle

C'est avec cette idée de représenter la conversation sous une forme intrinsèque à sa nature temporelle que je décide de reprendre une recherche d'inspirations.

**« La mélodicité, la haute régularité de la courbe mélodique, est étroitement liée à l'expression des émotions tendres. »**

C'est la notion de mélodicité de la parole humaine exposée par Iván Fónagy qui me conduit à regarder des travaux d'artistes et de designers menés notamment en lien avec la musique. Tout comme la conversation, la musique est temporelle et inclut parfois l'idée d'échanges entre deux ou plusieurs instruments, tels les *jam sessions* dans le jazz (fig. 3.7)



De toutes ces représentations, je retiens l'idée que la parole ou la note de musique s'inscrit sur une durée à un instant *t*. Le temps *y* est souvent matérialisé par une ligne directrice, horizontale ou non. Cette ligne permet donc d'orchestrer temporellement les notes ou dans la conversation, une suite d'idées et d'échanges, mais aussi de mettre en lumière les pauses et les silences. Quant à la durée, une des parfaites représentations est la note de musique qui est codifiée graphiquement pour illustrer sa valeur. Ainsi, la tête de la note peut, en autres, être noire pour représenter 1 temps ou blanche pour 2 temps. Un autre exemple est le code Morse qui exploite le point et le tiret pour exprimer deux durées différentes. Ma recherche m'a conduit à voir d'autres travaux exploitant la taille d'un objet pour représenter la durée<sup>1</sup>.

La partition de musique (fig. 3.8) est aussi un objet remarquable pour la présence d'informations contextuelles telles que le titre de l'oeuvre, son auteur et les quelques consignes qui accompagnent certaines notes. Les liaisons qui relient une note à une autre sur la partition ne sont pas sans me rappeler les liens qui émanent entre certaines idées d'une conversation. Toute cette recherche me conduit à poursuivre mon projet.

Figure 3.7. Illustration  
Photographie de Francis Wolf (1957) prise lors d'une session d'enregistrement de John Coltrane et Lee Morgan.

<sup>1</sup> Le projet *Off the Staff* de Nicholas Rougeux.

Figure 3.8. Illustration  
Partition *Clair de Lune* de Claude Debussy issue de la *Suite Bergamasque* (1890 - 1905)

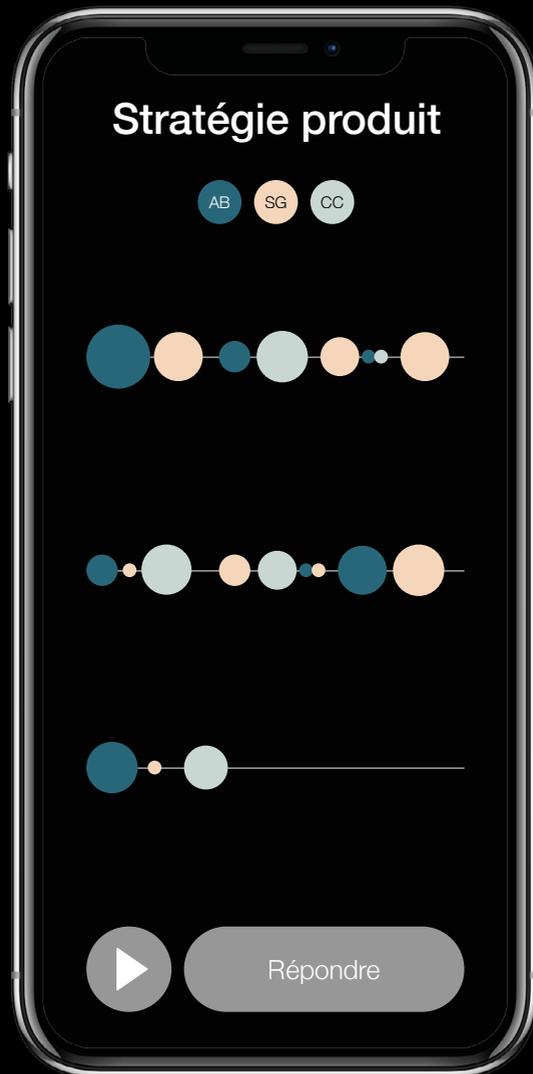


Figure 3.9. Illustration  
Esquisse d'Agoraphone représentant une conversation de trois professionnels sur le sujet de la stratégie produit.

### 3.6. Troisième version

Les recherches de représentations temporelles ont été source d'inspiration pour la poursuite du projet.

La nouvelle interface d'Agoraphone (fig. 3.9) intègre désormais une ligne de temps partant du haut, avec les enregistrements les plus anciens, pour aller vers le bas. De cette manière, l'architecture de l'information conserve la logique chronologique déjà présente dans les applications de messagerie instantanée (Messages, Whatsapp, Messenger, Slack, ...) où le message le plus récent est toujours présent en bas. Cette conception permet aussi d'exploiter le *scroll* et de pouvoir faire défiler la conversation vers le haut afin de retrouver un enregistrement plus ancien.

Cette ligne de temps est donc parsemée d'enregistrements audio représentés sous une forme sphérique, à l'image d'un point dans le temps. Parmi les qualités de cet objet, le diamètre représente la durée de l'enregistrement. Un point de petite taille est un enregistrement de courte durée et *vice versa*. Un algorithme peut être développé pour créer des points représentatifs de la durée en fonction d'un ratio durée (sec)/diamètre (px). La couleur du point représente le participant et permet de mettre en lumière graphiquement ceux et celles qui se sont exprimés et quand.

Issues de la partition de musique, l'interface intègre désormais quelques informations contextuelles que sont le titre de la conversation et les initiales des participants (fig. 3.9). Cette dernière information est issue de la deuxième version

d'Agoraphone (fig. 3.6) et de la volonté de représenter les interlocuteurs d'une conversation afin de créer un lien visuel. Un des chemins auxiliaires du travail de conception peut nous conduire à trouver des solutions pour donner accès visuellement à des informations complémentaires (nom, prénom, etc.) sur les personnes.

Concernant les fonctionnalités phares de l'application, l'interface incorpore un bouton de lecture de forme circulaire, en bas à gauche zone facilement accessible avec le pouce<sup>1</sup>. L'utilisateur peut aussi écouter un message en cliquant simplement dessus. Aussi ancré au bas de l'application, un bouton « Répondre » est présent permettant à l'utilisateur d'ajouter un nouveau message audio à la conversation.

Finalement, cette troisième version propose une approche basée sur la représentation temporelle qui semble plus naturelle que la représentation spatiale. Pour autant, l'idée d'Agora n'est pas abandonnée. La notion de lieu perdue à travers l'application et poursuit l'idée de se retrouver dans un lieu virtuel et commun pour échanger. Avec une représentation plus juste, l'application Agoraphone n'est pour autant pas viable sans la description du parcours utilisateur permettant d'ajouter un message. La suite du projet nous conduit donc à développer certaines fonctionnalités présentes sur l'interface graphique.

<sup>1</sup> Un article rédigé le 19 juillet 2016 par Joe Toscano pour Inside Design by Invision : Mobile first: Insights from going mobile only

## 3.7. Orchestrer la conversation

Évoqué dans la section 3.4, répondre dans une conversation n'est pas une action anodine. Le processus de projet ne nous conduit pas uniquement à représenter la conversation, il est aussi nécessaire de l'organiser. La question se pose car sans vue directe sur les autres participants, le dialogue peut vite se transformer en cacophonie et faire perdre le fil des idées et de la conversation.

Ainsi comment concevoir l'orchestration d'une conversation ? Est-il nécessaire que l'initiateur de celle-ci soit aussi le modérateur avec pour responsabilité la distribution de la parole ? Ces questions me conduisent à observer les conversations de la vie courante qui, en de rares occasions, sont dirigées par des modérateurs. Nos échanges quotidiens semblent être cadrés par des mécanismes conversationnels implicites. Je m'appuie alors sur le travail de Catherine Kerbrat-Orecchioni, professeure en science du langage et directrice de l'équipe « Groupe de Recherches sur les Interactions Communicatives » au CNRS.

Dans son livre *Les interactions verbales* [Kerbrat-Orecchioni, 1997], la professeure réalise une synthèse des travaux de recherches liés à l'analyse conversationnelle et l'étude des différents types d'interactions. La professeure caractérise la conversation comme des interlocuteurs qui parlent à tour de rôle, introduisant de ce fait le concept de « tour de parole ». Celui-ci repose sur un fondement : tout un chacun a le droit de s'exprimer et le devoir d'écouter les autres participants s'exprimer. Pour illustrer ce système, la professeure nous propose le cas suivant :

**« 1 - Le locuteur en place (L1 : « current speaker ») a le droit de garder la parole un certain temps, mais aussi le devoir de la céder à un moment donné ;**

**2 - son successeur potentiel (L2 : « next speaker ») a le devoir de laisser parler L1, et de l'écouter pendant qu'il parle ; il a aussi le droit de réclamer la parole au bout d'un certain temps, et le devoir de la prendre quand L1 la lui cède. »**

Ce système de droit et de devoir met en lumière que si une discussion entre deux protagonistes donne généralement un schéma de conversation simple et binaire : abababa, celui-ci présente aucune logique dès lors que le nombre de participants augmente : abcbaca ; acbabab ; etc. Toutefois, quel que soit le nombre de participants, l'auteur évoque des règles communes et fondamentales :

**« (1) - La fonction locutrice doit être occupée successivement par différents acteurs.**

**(2) - Une seule personne parle à la fois.**

**(3) - Il y a toujours une personne qui parle, c'est à dire que le temps de la conversation est pour l'essentiel occupé par la parole (le « speech stream » est quasiment continu), et que les intervalles (« gaps ») séparant les tours sont aussi réduits au minimum. »**

Ce système implicite lève le voile sur l'orchestration d'une conversation exemptée de modérateur. Toutefois, il reste à comprendre comment s'articule le passage d'un locuteur à un autre. La professeure est formelle: ce sont les participants eux-mêmes qui négocient les tours de paroles à l'aide d'une multitude de mécanismes. Dans un premier temps, c'est le locuteur qui, à l'approche de la fin de son tour, envoie « des signaux de fin de tour » de différentes natures : mimo-gestuelle, verbale, phonétique et prosodique [Annexe 3.1]. En réponse à ses signaux, un des protagonistes peut s'auto-sélectionner comme prochain locuteur, tout comme l'actuel locuteur peut désigner un des interlocuteurs avant de libérer la parole [Annexe 3.2].

Finalement, des recherches réalisées par Catherine Kerbrat-Orecchioni, nous retiendrons, pour la poursuite de la conception d'Agoraphone que le concept de « tour de parole » vient confirmer l'idée qu'une conversation est composée d'interventions successives. Ainsi, les tours de parole sont représentés par les points de différentes tailles sur la ligne de temps (section 3.6). Concernant l'orchestration de la conversation, « les signaux de fin de tour » présents dans les conversations de la vie courante ne peuvent être exploités pour le concept d'Agoraphone. Dans sa conception, la fin d'un tour de parole est signalé par la mise en ligne d'un enregistrement. De cette question d'orchestration, nous retiendrons des éléments qui permettent le choix du locuteur suivant. Ainsi le locuteur en place peut vouloir

désigner l'interlocuteur qui lui semble le plus approprié pour compléter la discussion sur une ou plusieurs idées. Agoraphone doit donc permettre au locuteur de désigner une personne à la fin de son tour. Si celui-ci ne désigne personne, un ou plusieurs interlocuteurs peuvent souhaiter prendre le tour de parole suivant par une manifestation. Dans la vie courante, les interlocuteurs peuvent prendre une posture d'appel, lever la main ou réaliser un signe pour montrer leur volonté de s'exprimer. Agoraphone doit permettre aux interlocuteurs de demander la parole pour le prochain tour. Se pose alors une dernière question, comment désigner le prochain locuteur si plusieurs se présentent au poste ? Les réponses semblent multiples et liées à la situation de la conversation. Toutefois, Catherine Kerbrat-Orecchioni nous donne une des réponses possibles :

**« idéalement, une conversation se caractérise aussi par :**

**1 - un équilibre relatif de la longueur des tours : plus long est le temps de parole, plus long sera le temps d'écoute d'un même locuteur, c'est-à-dire que les temps de parole sont positivement corrélés entre les locuteurs ; »**

### 3.8. Parcours utilisateur

Les travaux de recherches menés par Catherine Kerbrat-Orecchioni me permettent de décrire un premier parcours utilisateur qui s'appuie sur certains mécanismes conversationnels exposés par la professeure. Ainsi, le travail de conception se poursuit sur la base de l'interface présentée dans la troisième version (section 3.6). Néanmoins, les maquettes d'Agoraphone exposées dans les pages qui suivent (fig. 3.10 à 3.15) apportent leurs lots de nouveautés.

Désormais, la représentation graphique des personnes, signifiée par les initiales, évolue dans le temps et symbolise en même temps un classement lié à l'équilibrage de la parole (fig 3.10 : \*). Celui-ci permet d'évoquer le rang de chaque interlocuteur concernant son éligibilité à obtenir la parole au tour suivant. De ce fait, la personne la plus à gauche obtiendra en priorité la parole si elle en fait la demande. Pour exemple, entre le début (fig. 3.10) et la fin (fig. 3.15) de ce parcours illustré, le classement évolue.

Toujours placé en dernier sur la ligne chronologique, un nouveau type de point a aussi été conçu (fig 3.10 : \*\*). Celui-ci apparaît, à l'instant t, lorsque la parole est détenue par un des interlocuteurs. Sa forme reprend le symbole de l'enregistrement et est actualisée par la couleur du locuteur.

Les fonctionnalités ancrées en bas de l'application ont aussi été repensées. Alors que le *play* a été recentré, le bouton « Répondre » a été remplacé par un nouveau de forme circulaire et composé d'un symbole représentant un micro (fig. 3.10). Ce changement est issu des travaux de Catherine Kerbrat-Orecchioni qui évoquent le fait qu'avant même de répondre, les interlocuteurs qui souhaitent prendre la parole (ou le mi-

cro) se manifestent auprès du groupe. Ce bouton permet donc d'exprimer ce souhait.

Ce parcours utilisateur présente aussi l'interface d'enregistrement (fig 3.13) qui reprend les informations ancrées en haut de l'application (nom de la conversation et interlocuteurs). Au centre, le bouton d'enregistrement qui est représenté par son symbole évoqué plus tôt. D'une simple pression, l'utilisateur démarre l'enregistrement et peut livrer son message. Une nouvelle pression sur le bouton, l'arrêtera.

Dès lors que l'enregistrement est terminé et validé par l'utilisateur, l'application offre la possibilité à celui-ci de désigner le prochain locuteur (fig 3.14) et par la suite de publier son enregistrement.

Le parcours utilisateur présenté au travers des écrans qui suivent (fig. 3.10 à 3.15) met en scène trois protagonistes (AB, SG, CC), travaillant au sein d'une start-up et qui exposent tour à tour sur Agoraphone leurs idées concernant la stratégie du produit. Dans le premier écran (fig 3.10), AB, qui souhaite prendre la parole, constate que celle-ci est prise par CC. Toutefois, celui-ci est premier dans le classement pour le prochain tour de parole et appuie sur le bouton « micro » pour manifester son souhait de parler (fig 3.11). Dès lors que CC met à disposition son enregistrement, l'application oblige AB à écouter cette nouvelle intervention avant d'enregistrer la sienne. Ainsi, ce n'est que lorsque l'écoute est terminée que le bouton « micro » s'active en rouge pour signifier la possibilité d'enregistrer (fig 3.12). AB clique sur le bouton « micro », réalise son enregistrement (fig 3.13) qu'il peut réécouter avant de publier. Par la suite, AB choisit de donner la parole à SG (fig 3.14) et met en ligne son message (fig 3.15).

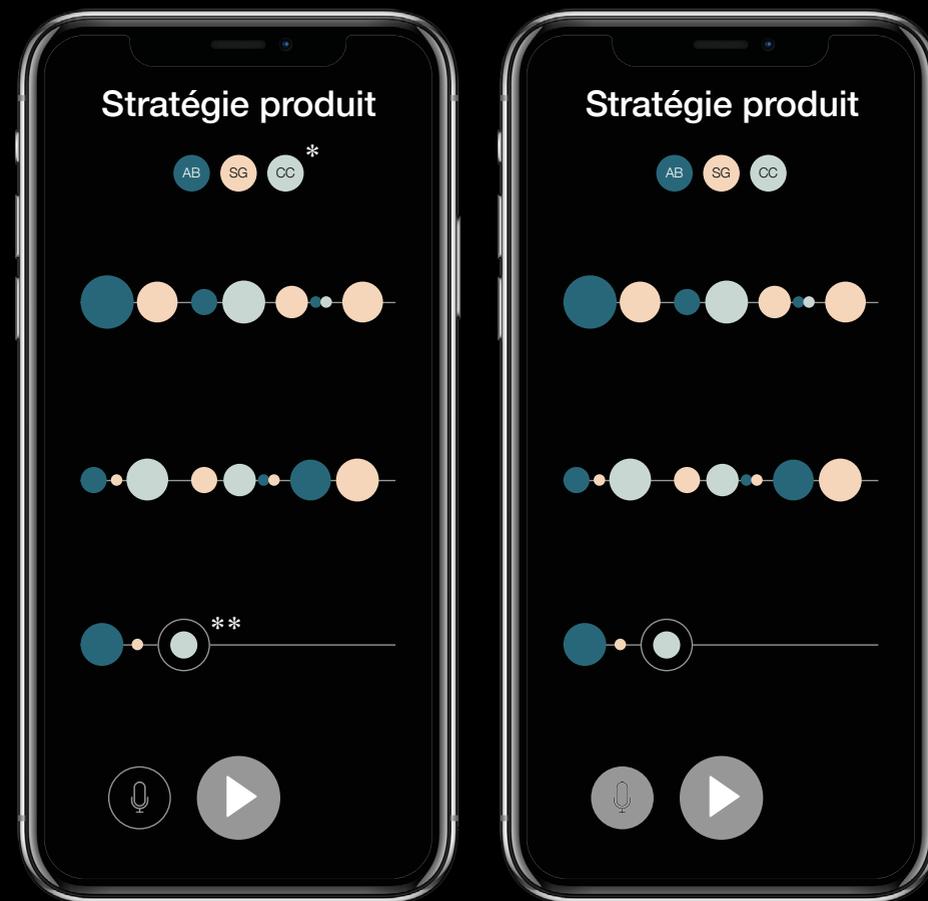


Figure 3.10. Illustration (à gauche) Esquisse d'Agoraphone représentant l'étape 01 du parcours où l'interlocuteur CC détient la parole.

Figure 3.11. Illustration (à droite) Esquisse d'Agoraphone représentant l'étape 02 du parcours où le souhait de parler est exprimé.

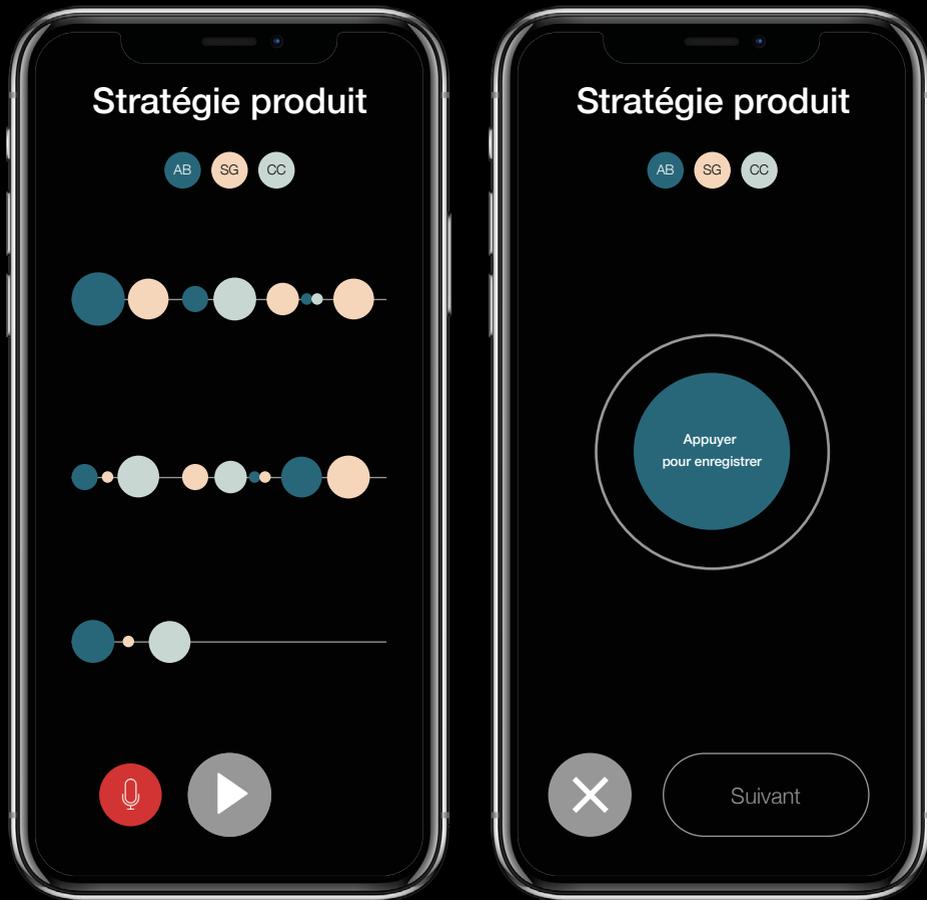


Figure 3.12. Illustration (à gauche)  
Esquisse d'Agoraphone représentant l'étape 03 du parcours où l'application indique la possibilité d'enregistrer

Figure 3.13. Illustration (à droite)  
Esquisse d'Agoraphone représentant l'étape 04 du parcours où l'interface permet l'enregistrement du message.

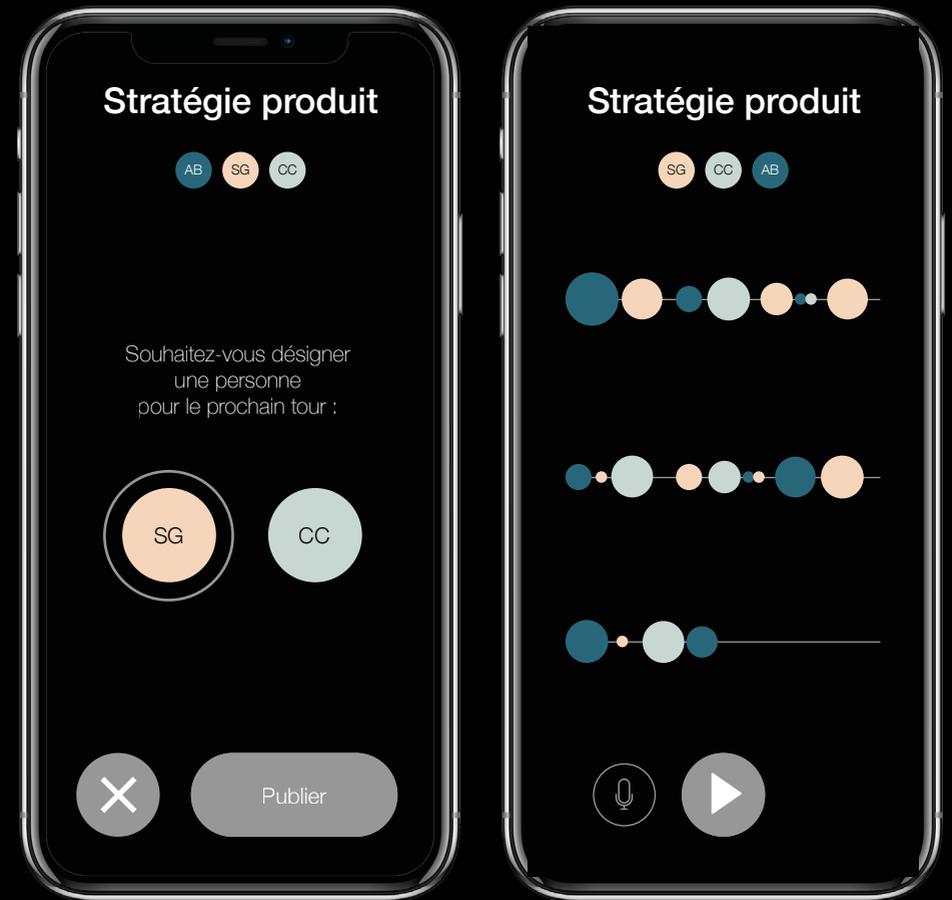


Figure 3.14. Illustration (à gauche)  
Esquisse d'Agoraphone représentant l'étape 05 du parcours où l'application permet de désigner le prochain locuteur.

Figure 3.15. Illustration (à droite)  
Esquisse d'Agoraphone représentant l'étape 06 du parcours présentant l'état de la conversation à la fin du parcours.

### 3.9. Et maintenant ?

À ce stade de la démarche, le travail de conception nous mène vers une première version « viable » de l'application Agoraphone. En tant que futur expert en stratégie d'innovation par le design, il devient nécessaire de valider l'intérêt et le potentiel de ce nouvel usage de communication vocale. Le test utilisateur de cet objet peut prendre de multiples formes. L'intérêt d'Agoraphone par le public peut être confirmé ou infirmé par la mise en ligne d'une *landing page*. Celle-ci peut, par le biais de textes explicatifs, de présentations de parcours utilisateur, de vidéo, mais aussi par des contenus audio, présenter le concept d'Agoraphone. Le visiteur de la page serait alors invité à s'inscrire pour télécharger l'application. Ainsi, l'intérêt porté pour l'objet serait évalué par le ratio : nombre d'inscrits sur nombre de visiteurs. Néanmoins, ce type de test ne propose que des données quantitatives qui nécessite un nombre important d'individus pour être validé statistiquement. Cette même *landing page* ou le parcours utilisateur peuvent aussi être présentés à des personnes lors d'entretiens. L'idée étant de recueillir des données qualitatives sur l'objet qui permettront également de valider la désirabilité du produit, mais aussi d'obtenir des informations sur l'usage et les éventuels contextes d'utilisation d'Agoraphone. Ces entretiens peuvent être complétés par la présentation d'un prototype que les personnes interrogées pourront tester. Ce second test peut permettre aux futurs utilisateurs de se projeter pleinement dans l'utilisation d'Agoraphone et d'énoncer des informations complémentaires. L'ensemble des informations recueillies devront être analysées afin de déterminer la direction du processus de conception et assurer la viabilité du projet.

Ce processus de conception peut aussi se poursuivre par des axes de travail révélés à partir de problèmes présents dans des applications exploitant la parole humaine. Ainsi, la retranscription graphique de certaines informations et contenus présents dans les enregistrements audio semblent primordial pour faciliter l'utilisation en tout temps de l'application Agoraphone. Ce problème<sup>1</sup> est notamment mis en évidence par l'usage du voice messaging via les applications de messagerie instantanée. Toutefois, l'idée n'est pas de copier sans réflexion la stratégie énoncée par ces entreprises. Il convient donc de questionner le problème lié à l'accès au contenu et la possibilité d'en obtenir rapidement la teneur. Ceci n'est pas sans rappeler les emails dont le contenu est masqué lorsque nous sommes à l'entrée de la boîte. Pour autant, nous connaissons la teneur ou le sujet de celui-ci à travers l'objet. Ainsi de nouvelles possibilités se dessinent pour Agoraphone comme le résumé sonore présent dans la musique. D'autre part, les enregistrements sonores étant les vrais objets de l'application Agoraphone, il sera nécessaire de penser des interfaces qui soutiennent l'écoute et de comprendre comment les utilisateurs écoutent une conversation (écoute totale, partielle, accélérée) pour développer un lecteur audio adapté.

Les axes de conception peuvent aussi être guidés par des composants présents dans les conversations de la vie réelle. À ce sujet, Catherine Kerbrat-Orecchioni évoque les signaux d'écoute (fig 3.16) présents dans nos échanges [Kerbrat-Orecchioni, 1997] et dont le développement est à évaluer :

**« De son côté, le récepteur doit lui aussi produire des signaux, que nous dirons régulateurs (ou « signaux d'écoute »), et qui apparaissent dans la littérature sous des habillages divers : procédés de « feed-back » ou de « back-channel », [...]. Ces régulateurs ont des réalisations diverses, non verbales (regard, hochement de tête, froncement des sourcils, bref sourire, léger changement postural), vocales (« mmh » et autres vocalisations), ou verbales (avec différents degrés d'élaboration : morphème exclamatif, ou à valeur d'approbation - « oui », « d'accord » -, reprise avec ou sans reformulation, par L2 cette fois, du discours de L1). »**

En dernier lieu, les axes peuvent aussi être induits par des inspirations issues de la recherche menée en amont de la conception. Dans le cas d'Agoraphone, ce sont notamment les liaisons entre les notes d'une partition de musique (fig. 3.8) qui m'interpellent. Évoqué précédemment (section 3.5), ceci n'est pas sans me rappeler les diverses idées qui se lient et s'agrémentent au cours d'une conversation. Afin de faciliter la lecture d'une conversation ou de cibler le développement d'une idée, la réalisation d'une liaison graphique et auditive pourrait être un axe à développer.

<sup>1</sup> Un article rédigé le 2 décembre 2019 par Chris Stokel-Walker pour *The Guardian* : *Voice messaging – conversational gain or pain?*

Figure 3.16. Illustration Photographie prise par Eric Cabanis lors d'une manifestation de *Nuit debout* où des participants expriment leur accord à l'aide d'un mouvement particulier des mains





## Conclusion

---

À l'origine de ce mémoire, c'est le constat d'intermédiation de la technologie numérique dans les relations humaines qui a initié la recherche d'une communication entre pairs plus directe et plus humaine, dans ce même numérique. La problématique posée en introduction nous a offert un travail en trois temps.

Un premier chapitre consacré à la parole humaine a permis de confirmer que ce mode de communication expose une richesse d'informations, amputé à l'écrit. La parole humaine, au travers de la voix, est vive, expressive, emplit d'émotions. Pour le récepteur, ces informations sont notamment perçues par la mélodie de la parole. Il ne fait nul doute que l'être humain peut identifier, éprouver ce que ses pairs ressentent et faire preuve d'empathie. La voix et la parole sont aussi une empreinte vocale, propre à chaque être humain et détenant des informations que nous sommes tous à même de décoder (genre, âge, origine géographique, etc.).

Dans le deuxième chapitre, l'engouement pour la parole humaine nous a conduit à porter un regard

sur cet objet insaisissable dans les technologies passées et actuelles. La découverte fut le retour d'objets liés à la parole humaine, initié par l'arrivée du smartphone en 2007. L'instrumentalisation de la parole nous conditionne désormais à employer ce mode de communication dans nos interactions avec la machine, mais aussi entre êtres humains. Cette tendance nommée *voice messaging* est notamment motivée par l'envie de transmettre ses émotions à ses pairs. Cette situation semble donc créer un *momentum* favorisant le développement d'usages liés à la parole et nous incite dans un dernier chapitre à développer un processus de projet en réponse à la problématique.

La démarche de projet, basée sur une intuition et non sur la résolution d'un problème vécu, nous mène dans un cheminement de création parsemé d'inspirations, de recherches et d'actes de conception. L'objet développé, baptisé Agoraphone, se définit comme un lieu virtuel pour se retrouver et échanger vocalement à l'aide d'enregistrements audio.

Ainsi, Agoraphone se pose comme une réponse à la problématique, soit un outil numérique pour favoriser des interactions humaines plus empathiques. Néanmoins, la capacité de l'application à atteindre cet objectif n'a pas été évaluée dans ce mémoire.

D'autre part, la démarche de processus de création porte un bénéfice majeur : la possibilité de se projeter dans de futurs usages numériques et d'en imaginer de nouveaux. Au cours du chapitre trois, la possibilité d'exploiter Agoraphone comme un outil de travail pour les professionnels travaillant à distance a été évoquée, mais n'est pas unique. Entre autres, l'objet peut aisément

être utilisé par le grand public dans le cadre de conversations personnelles, pour des débats publics, par une seule et même personne pour conserver des pensées ou encore afin de présenter oralement son parcours professionnel lors de la candidature à un futur emploi, etc.

Finalement, formé à résoudre des problèmes, ce mémoire aura été pour moi une opportunité d'apprendre à travailler et développer une intuition. Par cet exemple, j'y ai acquis une méthodologie de travail reposant sur l'acte de questionner, la recherche et le processus de création. En soi, une méthodologie nécessaire à l'acte d'innover.

## Bibliographie

---

ABITBOL, Jean. L'odyssée de la voix. Robert Laffont, 2005.

BOLT, Richard A. "Put-that-there": Voice and gesture at the graphics interface. ACM, 1980.

BREGUET, Antoine. La transmission de la parole : le phonographe, le microphone, l'aérophone. Revue des Deux Mondes (1829-1971), 1878, vol. 28, no 3, p. 689-708.

COCTEAU, Jean. La voix humaine. Stock, 2001.

EDISON, Thomas A. The phonograph and its future. The North American Review, 1878, vol. 126, no 262, p. 527-536.

FÓNAGY, Ivan. La vive voix: essais de psycho-phonétique. Payot, 1983.

JONZE, Spike. her [Motion picture]. United States: Warner Bros, 2013.

KERBRAT-ORECCHION, Cathérine. Les Interactions verbales. vol. 1, approche interactionnelle et structure des conversations. Armand Colin. 1997.

KUBRICK, Stanley. 2001: a space odyssey [Motion picture]. Metro-Goldwyn-Mayer. 1968.

LUGER, Ewa & SELLEN, Abigail. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In : Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 2016. p. 5286-5297.

RÉVIS, Joana. La voix et soi: Ce que notre voix dit de nous. De Boeck Supérieur, 2013.

ROBINSON, Arthur L. More people are talking to computers as speech recognition enters the real world. Science, 1979, vol. 203, no 4381, p. 634-638

TURING, Alan M. Computing machinery and intelligence (1950). *The Essential Turing: The Ideas that Gave Birth to the Computer Age*. Ed. B. Jack Copeland. Oxford: Oxford UP, 2004, p. 433-64.

TURKLE, Sherry. *Alone together: Why we expect more from technology and less from each other*. 2011.

TYLDUM, Morten. *The Imitation Game* [Motion picture]. Black Bear Pictures, 2014.

VERLAINE, Paul. *La bonne chanson*. 1870.

WEIZENBAUM, Joseph, et al. ELIZA--A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 1966, vol. 9, no 1, p. 36-45.

## Annexes

—

## Annexe 1.1

Extrait du livre *La vive voix* d' Iván Fónagy (p.34) qui décrit l'influence des émotions sur le corps :

« On arrive sans peine à situer tous les traits vocaux de la colère ou de la haine dans les cadres du combat ancestral. La tension musculaire pourrait correspondre à la préparation à la lutte ; ou préfigurer le combat lui-même. La grande distance maxillaire, la bouche grande ouverte constituent une menace directe de l'adresse de l'ennemi, ainsi que la mâchoire supérieure avancée avec les dents qui mordent sur la lèvre inférieure. Le timbre sombre rappelle la voix des fauves féroces, c'est également celles des mâles robustes. Le roulement très fort des /r/ apical peut être conçu en termes de menace, mais aussi comme une performance symbolique du combat. »

## Annexe 1.2

Extrait du livre *La vive voix* d' Iván Fónagy (p.31) qui décrit les signes physiologiques de la colère dans la zone pharyngo-buccale :

« Le colère se manifeste en français comme en hongrois :

- a) par des mouvements saccadés de la langue : des transitions extrêmement rapides sont suivies de périodes de figement prolongeant la période de la tenue dans les positions extrêmes (la position la plus fermée pour les voyelles fermées, la plus ouverte pour les voyelles ouvertes) ;
- b) par la tension musculaire linguale, labiale, pharyngée renforcée : la langue est fortement pressée contre les alvéoles pour le /t/, contre le palais pour le /k/ ; la superficie du contact est, par conséquent, souvent redoublée par rapport à l'articulation dans la parole neutre ;
- c) Les voyelles /i/, /y/, /u/ fermées sont plus fermées, les voyelles ouvertes nettement plus ouvertes ; ainsi l'angle maxillaire et la distance entre les lèvres accroît dans le cas de /a/ de 15 mm à 25 mm ;
- d) La langue est plus rétractée pour les voyelles comme pour les consonnes par rapport à la parole neutre, ce qui prête aux sons un timbre plus sombre ;
- e) La mandibule (la mâchoire inférieure) est rétractée, les incisives supérieures s'avancent, ce qui fait que les incisives supérieures « mordent » sur la lèvre inférieure en prononçant les semi-voyelles /w/. »

Extrait du livre *La vive voix* d' Iván Fónagy (p.31) qui décrit les signes physiologiques de la colère dans la zone glottale :

« Dans la colère rentrée (voix étranglée), les cordes vocales sont fortement accolées, contractées, les bandes ventriculaires s'imbriquent, les ventricules Morgagni sont encore plus rétrécis, et restent souvent invisibles. La glotte se transforme en un canal extrêmement étroit. L'espace super-glottique se rétrécit en même temps. Les analyses radiographiques de Luchsinger et du Troyan mettent en évidence le rapport entre le rétrécissement laryngé et pharyngé. »

## Annexe 2.1

Extrait de l'article *The phonograph and its future* publié dans *The north american review* et écrit par de Thomas Edison qui décrit les usages possibles du phonographe :

« The probable application of these properties of the phonograph and the various branches of commercial and scientific industry presently indicated will require the exercise of more or less mechanical ingenuity. Conceding that the apparatus is practically perfected in so far as the faithful reproduction of sound is concerned, many of the following applications will be made the moment the new form of apparatus, which the writer is now about completing, is finished. These, then, might be classed as actualities; but they so closely trench upon other applications which will immediately follow, that it is impossible to separate them: hence they are all enumerated under the head of probabilities, and each specially considered. Among the more important may be mentioned: Letter-writing, and other forms of dictation books, education, reader, music, family record; and such electrotype applications as books, musical-boxes, toys, clocks, advertising and signaling apparatus, speeches, etc., etc. Letter-writing. The apparatus now being perfected in mechanical details will be the standard phonograph, and may be used for all purposes, except such as require special form of matrix, such as toys, clocks, etc., for an indefinite repetition of the same thing. The main utility of the phonograph, however, being for the purpose of letter-writing and other forms of dictation, the design is made with a view to its utility for that purpose. The general principles of construction

are, a flat plate or disk, with spiral groove on the face, operated by clock-work underneath the plate; the grooves are cut very closely together, so as to give a great total length to each inch of surface? a close calculation gives as the capacity of each sheet of foil, upon which the record is had, in the neighborhood of 40,000 words. The sheets being but ten inches square, the cost is so trifling that but 100 words might be put upon a single sheet economically. Still, it is problematical whether a less number of grooves per inch might not be the better plan? it certainly would for letters? but it is desirable to have but one class of machine throughout the world; and as very extended communications, if put upon one sheet, could be transported more economically than upon two, it is important that each sheet be given as great capacity as possible. The writer has not yet decided this point, but will experiment with a view of ascertaining the best mean capacity. The practical application of this form of phonograph for communications is very simple. A sheet of foil is placed in the phonograph, the clock-work set in motion, and the matter dictated into the mouth-piece without other effort than when dictating to a stenographer. It is then removed, placed in a suitable form of envelope, and sent through the ordinary channels to the correspondent for whom designed. He, placing it upon his phonograph, starts his clock-work and listens to what his correspondent has to say. Inasmuch as it gives the tone of voice of his correspondent, it is identified. As it may be filed away as other letters, and at any subsequent time reproduced, it is a perfect record. As two sheets of foil have been indented with the same facility as a single sheet, the « writer « may thus keep a du-

PLICATE of his communication. As the principal of a business house, or his partners now dictate the important business communications to clerks, to be written out, they are required to do no more\* by the phonographic method, and do thereby dispense with the clerk, and maintain perfect privacy in their communications. The phonograph letters may be dictated at home, or in the office of a friend, the presence of a stenographer not being required. The dictation may be as rapid as the thoughts can be formed, or the lips utter them. The recipient may listen to his letters being read at a rate of from 150 to 200 words per minute, and at the same time busy himself about other matters. Interjections, explanations, emphasis, exclamations, etc., may be thrown into such letters, ad libitum. In the early days of the phonograph, ere it has become universally adopted, a correspondent in Hong-Kong may possibly not be supplied with an apparatus, thus necessitating a written letter of the old-fashioned sort. In that case the writer would use his phonograph simply as a dictating-machine, his clerk writing it out from the phonograph at leisure, causing as many words to be uttered at one time as his memory was capable of retaining until he had written them down. This clerk need not be a stenographer, nor need he have been present when the letter was dictated, etc. The advantages of such an innovation upon the present slow, tedious, and costly methods are too numerous, and too readily suggest themselves, to warrant their enumeration, while there are no disadvantages which will not disappear coincident with the general introduction of the new method. Dictation. All kinds and manner of dictation which will permit of the application of the mouth

of the speaker to the mouth-piece of the phonograph may be as readily effected by the phonograph as in the case of letters. If the matter is for the printer, he would much prefer, in setting it up in type, to use his ears in lieu of his eyes. He has other use for them. It would be even worth while to compel witnesses in court to speak directly into the phonograph, in order to thus obtain an unimpeachable record of their testimony. The increased delicacy of the phonograph, which is in the near future, will enlarge this field rapidly. It may then include all the sayings of not only the witness, but the judge and the counsel. It will then also comprehend the utterances of public speakers. Books may be read by the charitably-inclined professional reader, or by such readers especially employed for that purpose, and the record of such book used in the asylums of the blind, hospitals, the sick-chamber, or even with great profit and amusement by the lady or gentleman whose eyes and hands may be otherwise employed; or, again, because of the greater enjoyment to be had from a book when read by an elocutionist than when read by the average reader. The ordinary record-sheet, repeating this book from fifty to a hundred times as it will, would command a price that would pay the original reader well for the slightly-increased difficulty in reading it aloud in the phonograph. Educational Purposes. As an elocutionary teacher, or as a primary teacher for children, it will certainly be invaluable. By its difficult passages may be correctly rendered for the pupil but once, after which he has only to apply to his phonograph for instructions. The child may thus learn to spell, commit to memory, a lesson set for it, etc., etc. Music. The phonograph will undoubtedly be liberally de

voted to music. A song sung on the phonograph is reproduced with marvelous accuracy and power. Thus a friend may in a morning-call sing us a song which shall delight an evening company, etc. As a musical teacher it will be used to enable one to master a new air, the child to form its first songs, or to sing him to sleep. Family Record. For the purpose of preserving the sayings, the voices, and the last words of the dying member of the family as of great men the phonograph will unquestionably outrank the photograph. In the field of multiplication of original matrices, and the indefinite repetition of one and the same thing, the successful electrotyping of the original record is an essential. As this is a problem easy of solution, it properly ranks among the probabilities. It comprehends a vast field. The principal application of the phonograph in this direction is in the production of Phonographic Books. A book of 40,000 words upon a single metal plate ten inches square thus becomes a strong probability. The advantages of such books over those printed are too readily seen to need mention. Such books would be listened to where now none are read. They would preserve more than the mental emanations of the brain of the author; and, as a bequest to future generations, they would be unequalled. For the preservation of languages they would be invaluable. Musical-Boxes, Toys, etc. The only element not absolutely assured, in the result of experiments thus far made which stands in the way of a perfect reproduction at will of Adelina Patti's voice in all its purity is the single one of quality, and even that is not totally lacking, and will doubtless be wholly attained. If, however, it should not, the musical-box, or cabinet, of the present, will be superseded by that which

will give the voice and the words of the human songstress. Toys. A doll which may speak, sing, cry, or laugh, may be safely promised our children for the Christmas holidays ensuing. Every species of animal or mechanical toy such as locomotives, etc. may be supplied with their natural and characteristic sounds. Clocks. The phonographic clock will tell you the hour of the day; call you to lunch; send your lover home at ten, etc. Advertising, etc. This class of phonographic work is so akin to the foregoing, that it is only necessary to call attention to it. Speech and other Utterances. It will henceforth be possible to preserve for future generations the voices as well as the words of our Washingtons, our Lincolns, our Gladstones, etc., and to have them give us their «greatest effort» in every town and hamlet in the country, upon our holidays. Lastly, and in quite another direction, the phonograph will perfect the telephone, and revolutionize present systems of telegraphy. That useful invention is now restricted in its field of operation by reason of the fact that it is a means of communication which leaves no record of its transactions, thus restricting its use to simple conversational chit-chat, and such unimportant details of business as are not considered of sufficient importance to record. Were this different, and our telephone-conversation automatically recorded, we should find the reverse of the present status of the telephone. It would be expressly resorted to as a means of perfect record. In writing our agreements we incorporate in the writing the summing up of our understanding using entirely new and different phraseology from that which we used to express our understanding of the transaction in its discussion, and not infrequently thus begetting

perfectly innocent causes of misunderstanding. Now, if the telephone, with the phonograph to record its sayings, were used in the preliminary discussion, we would not only have the full and correct text, but every word of the whole matter capable of throwing light upon the subject. Thus it would seem clear that the men would find it more advantageous to actually separate a half-mile or so in order to discuss important business matters, than to discuss them verbally, and then make an awkward attempt to clothe their understanding in a new language. The logic which applies to transactions between two individuals in the same office, applies with the greater force to two at a distance who must discuss the matter between them by the telegraph or mail. And this latter case, in turn, is reinforced by the demands of an economy of time »

## Annexe 3.1

Extrait du livre *Les interactions verbales* de Catherine Kerbrat-Orecchioni (p.166) qui décrit les signaux de fin de tour dans la conversation :

« Cette place transitionnelle, L1 doit l'inscrire d'une certaine manière dans l'énoncé, à l'aide de l'un et/ou l'autre des « signaux de fin de tour » que permet le système.

Ces signaux peuvent être regroupés en trois catégories, distinguées sur la base de leur statut sémiotique :

(1) Signaux de nature verbale :

. Complétude syntactico-sémantique de l'énoncé - mais cette propriété ne constitue une condition ni suffisante, ni même nécessaire (car il est commun que l'on cesse de parler sans savoir achevé sa phrase), pour qu'il y ait fin de tour.

. Statut illocutoire de l'énoncé : certains types d'actes (ex. : les questions) ont plus que d'autres vocation à fonctionner comme la première partie d'une paire adjacente, donc à solliciter un enchaînement immédiat - mais reste toujours, pour L1, la possibilité d'effectuer un rajout ou une expansion quelconque.

. Morphèmes connotant la clôture (« bon », « voilà », etc.) - mais ils n'impliquent pas nécessairement la clôture d'un tour.

. Plus explicites sont les expressions phatique (qui n'excluent toutefois pas elles non plus la possibilité d'une relance) telles que « hein ? », « non ? », voire « qu'est-ce que tu en penses ? » ; sans parler des formules métadiscursives dont l'usage reste très exceptionnel, du genre « j'ai fini », ou « à toi ! ».

(2) Signaux de nature phonétique, et surtout prosodique :

. Courbe intonative marquée (selon le type d'énoncé, nettement ascendante, ou descendante, sur les dernières syllabes).

. Pause de la voix - mais on a vu que le « gap » minimal était de trop courte durée pour être perceptible, et qu'on rencontrait des pauses à l'intérieur des tours. Pour fonctionner comme un signal efficace de fin de tour, la pause doit en fait être annoncée par

. un ralentissement du débit, une chute d'intensité articulatoire, une voix traînante sur les dernières syllabes de l'énoncé ; et s'accompagner de certains indices non verbaux.

(3) Signaux de nature mimo-gestuelle :

. Regard soutenu porté en fin de tour sur le destinataire (alors qu'en cours de tour, L1 ne regarde L2 que par intermittence, histoire de vérifier que le contact est bien maintenu, et de capter les signes mimo-gestuels que produit l'interlocuteur) ; éventuellement, sollicitation plus nette de L2 par un mouvement approprié de la tête (levée du menton, ou simplement des sourcils).

. Achèvement de la gesticulation en cours, et relâchement général de la tension musculaire.»

## Annexe 3.2

Extrait du livre *Les interactions verbales* de Catherine Kerbrat-Orecchioni (p.169) qui décrit la nature du « next speaker » :

« Une fois que L1 a manifesté son désir d'abandonner la parole, quelqu'un doit la prendre, mais qui ? - le problème ne se posant évidemment que lorsque le groupe conversationnel comprend plus de deux participants ayant droit à la parole.

Sacks et al. distinguent deux techniques d'allocation du tour, et de sélection du « successeur » :

(1) C'est L1 qui sélectionne L2 (« next turn is allocated by current speaker selecting a new speaker »), à l'aide de certains procédés verbaux et/ou non verbaux.

1 - Procédés verbaux :

. Nomination explicite de L2 ; procédé très exceptionnel dans les conversations familières, mais qui se rencontre communément en situation scolaire (Sinclair et Coulthard 1975).

. Plus subtilement et sur le mode implicite, le contenu des propos de L1 peut jouer un certain rôle dans la sélection de L2 : s'ils « concernent » plus directement x que y, ou s'ils relèvent davantage de son domaine de compétence, x apparaîtra comme un successeur plus légitime que y (en vertu du « postulat selon lequel est censé apporter sa contribution au discours celui qui est tenu pour détenteur d'un savoir dans le domaine abordé par le locuteur » - Heddesheimer et Roussel 1986). Au niveau verbal, on a donc affaire à des procédés explicites mais rares, ou fréquents mais peu contraignants.

2 - Plus importants sont les procédés non verbaux

: geste d'ostension parfois, et surtout orientation du corps, et direction du regard, le participant sur lequel se stabilise en fin de tour le regard de L1 étant en principe intronisé successeur privilégié - et c'est bien ainsi, d'après les observations de Coulthard (1977), que les choses se passent en général.

(2) Auto-sélection : L2 se sélectionne lui-même comme « successeur » de L1. Après le « gap » de rigueur, L2 commence à parler : étant le « first starter », il devient le « next speaker ».

Les caractéristiques paraverbales des débuts de tour sont généralement les suivantes : registres plus aigu, intensité plus forte, débit accéléré, posture plus tendue, activité gestuelle plus intense - ces caractéristiques s'estompant progressivement à mesure que L2 se sent mieux assuré dans son statut de parleur (car ce n'est pas tout de prendre la parole : encore faut-il la garder, et la défendre contre d'éventuels rivaux).

1- Il peut en effet se faire qu'aucun candidat ne se présente : l'intervention de L1 est alors suivie d'un silence plus ou moins prolongé.

2- Il peut au contraire arriver que plusieurs candidats se présentent en même temps : il y a alors chevauchement, et pour remédier à cette espèce de carambolage verbal, les parties concernées doivent procéder à une négociation, qui peut s'effectuer sur un mode implicite ou explicite, courtois ou conflictuel (« pardon », « je vous en prie », vs « laissez-moi parler s'il te plaît », « c'est mon tour », etc.). »

