

PAROLE et GESTALTTHEORIE

Groupe de Recherche

M. E. LEIPP - Melle CASTELLENGO

M. J.S. LIENARD

LABORATOIRE D'ACOUSTIQUE DE LA FACULTE DES SCIENCES
8 Rue Cuvier - Paris 5°. Adresse postale 9 Quai St Bernard

P L A N

- 1°) GENERALITES
 - 2°) LE SYSTEME PHONATOIRE :
 - phénomènes quasi-stationnaires
 - phénomènes évolutifs
 - 3°) LA NATURE INFORMATIVE DU MESSAGE PARLE
 - 4°) EXTRACTION DE L'INFORMATION SEMANTIQUE
 - 5°) LES MOTS SONT DES FORMES :
 - la parole relève de la Gestalttheorie
 - 6°) LE DICTIONNAIRE DES ELEMENTS PHONETIQUES
 - 7°) LA SYNTHÈSE DE LA PAROLE PURE
-

PAROLE ET GESTALT-THEORIE

(Groupe de recherche sur la parole, du Laboratoire d'Acoustique de la Faculté des Sciences de Paris - Mécanique Physique)

M.E. LEIPP; Melle M. CASTELLENGO; M. J.S. LIENARD

I - GENERALITES

La parole pose des problèmes nombreux et compliqués, qui rentrent à peu près tous dans le cadre de nos recherches générales sur les signaux acoustiques, sur leur génération, leur structure physique, leur perception et leur intégration. Elle se distingue des autres messages acoustiques uniquement par la structure et le fonctionnement particulier de l'"instrument" qui sert à la fabriquer (le système phonatoire) et par la nature spéciale du message qu'elle véhicule (idées).

II - LE SYSTEME PHONATOIRE

Son anatomie est bien connue. On y distingue :

- une soufflerie, les poumons, fournissant l'énergie nécessaire sous forme d'air comprimé.
- une anche double, les cordes vocales, susceptible de moduler le débit aérien, donc de fabriquer des spectres de raies harmoniques réglables à volonté.
- une série de cavités dont les dimensions et les ouvertures peuvent être modifiées à volonté de façon autonome ou simultanée, à l'intérieur de positions-limites déterminées par l'anatomie et la physiologie des organes phonatoires (mâchoires, langue, lèvres, etc...)

Ces cavités jouent le rôle de résonateurs amortis couplés en série (V1, V2, V3, V4) ou en parallèle (V5) (fig.1).

...../

fig1

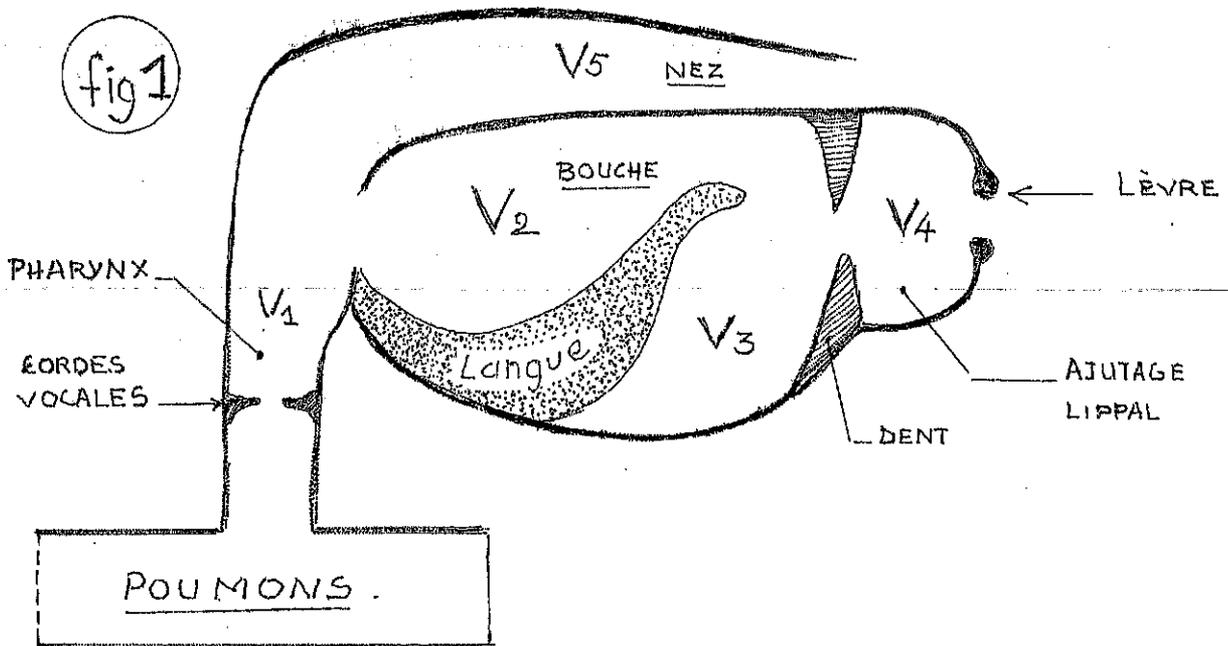


fig2

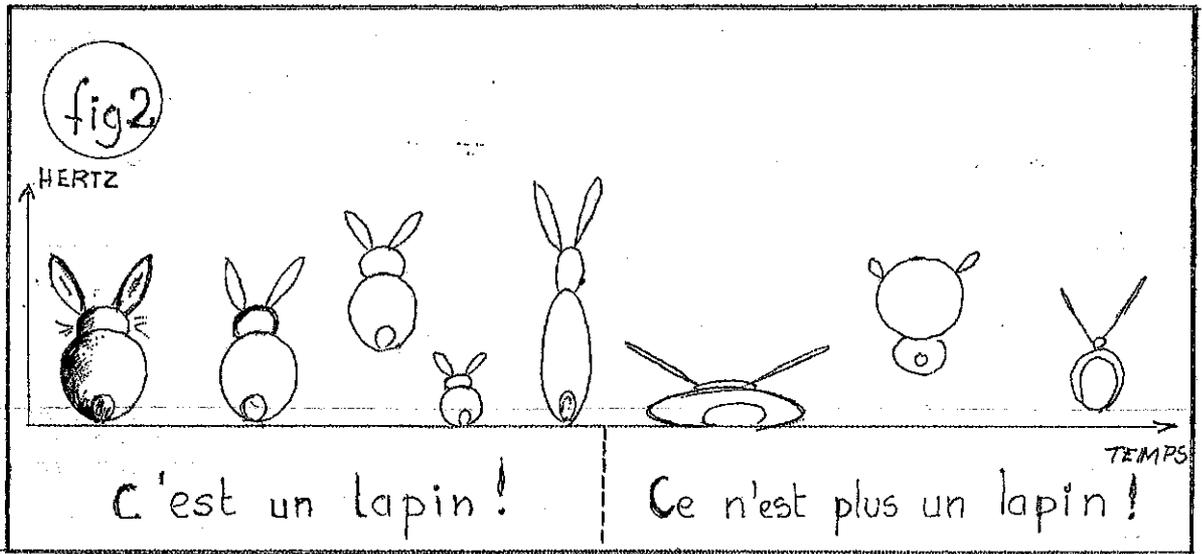
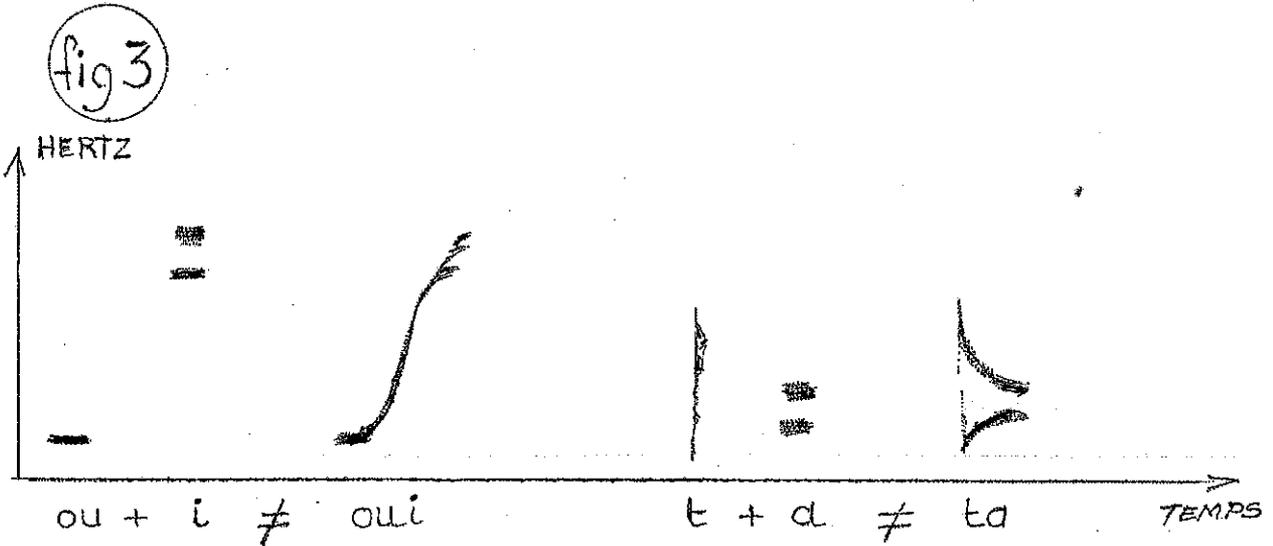


fig3



Les partiels 1 (fondamentaux) de ces résonateurs conditionnent les " formants ", bandes de fréquence dont la situation sur l'échelle des fréquences est fonction directe de l'ouverture du résonateur, et fonction inverse de son volume.

Dans la voix humaine normale, l'ensemble du système est excité par le spectre de raies délivré par les cordes vocales, et par le bruit d'écoulement du fluide aérien. Dans la parole chuchotée, les cordes vocales sont immobilisées.

Dans les deux cas, le système fournit soit :

- a) des phénomènes quasi-stationnaires lorsque les organes sont immobilisés dans une position donnée. On peut ainsi produire des " phonèmes " : voyelles, sifflantes, chuintantes, etc...) dont il est facile de connaître la structure spectrale à l'aide des méthodes d'analyse classiques.
- b) des phénomènes évolutifs. Lors de la parole normale, on observe facilement que le système phonatoire est en mouvement continu, sans arrêts notables sur des positions fixes. La parole n'est donc pas une succession de phonèmes accolés; la parole est mouvement. Si les organes changent continuellement de position, il est évident que le " spectre " change continuellement de composition; acoustiquement, la parole est un phénomène transitoire, une forme dynamique (alors que le phonème est une forme statique).

La seule représentation significative d'un phénomène acoustique évolutif est le sonagramme, qui donne l'évolution simultanée des trois dimensions d'une forme acoustique (niveau-fréquence-temps), donc un document permettant de saisir d'emblée, visuellement, un phénomène acoustique - perçu par ailleurs d'emblée par le système auditif. Moyennant un certain apprentissage, on réussit à faire une corrélation parfaite entre sonagramme et sensation auditive. Le sonagramme est maintenant bien connu de tous; nous l'utilisons constamment dans ce qui suit.

...../

III - LA NATURE INFORMATIVE DU MESSAGE PARLE

L'ensemble des sons complexes possibles avec le système phonatoire humain constitue la voix humaine. La voix ne devient parole que si elle véhicule une information intelligible, susceptible d'être captée et intégrée par un récepteur humain. Ce message est une idée, une forme dynamique mentale, que l'organe phonatoire transpose en forme dynamique acoustique suivant un code (la langue). Cette forme acoustique chemine dans divers canaux de communication où elle est plus ou moins déformée et amputée, mais normalement elle arrive au voisinage du récepteur sous forme reconnaissable si celui-ci possède le code. La parole humaine s'est élaborée à l'origine en fonction des conditions contraignantes de l'anatomie et de la physiologie humaines (émission et réception) et des conditions d'utilisation. Il fallait que le message " porte " à la plus grande distance possible et qu'il ne soit pas détruit par le bruit de fond ambiant, on devrait dire par les bruits de fond très variés qui nous entourent. L'expérience montre qu'elle est très efficace de ce point de vue; elle reste intelligible même si elle est considérablement altérée, déformée, détruite. Cela n'est possible que si elle possède une redondance énorme. Pour des raisons théoriques évidentes (étude de l'intelligibilité, étude des défauts de la parole, rééducation, synthèse, etc...) il est du plus haut intérêt de pouvoir éliminer expérimentalement la redondance et l'information parasitaire, en bref, d'extraire les invariants, la quintessence de la parole; de définir la parole à l'état pur, la forme essentielle qui véhicule l'idée, en un mot d'extraire l'information sémantique de la parole.

V - EXTRACTION DE L'INFORMATION SEMANTIQUE

Nous avons d'abord observé que dans la langue française (contrairement aux langues à tons : chinois, vietnamien etc...) le spectre de raies ne véhiculait pratiquement aucune information sémantique, mais uniquement de l'information esthétique. Le spectre de raies nous permet de reconnaître le locuteur (homme, femme, enfant, ami détermi-

né, etc...) en fonction de l'intensité, du rythme, de l'accent, du timbre de la voix, etc... Nous avons déjà montré à l'occasion d'un exposé au GALF (Un vocoder mécanique, la guimbarde; Annales Telecomm, Tome 18, 5-6-1963) que l'on pouvait fabriquer une parole semi-synthétique de la façon suivante :

On déconnecte soufflerie et cordes vocales; la bouche fait les mouvements normaux de prononciation d'un mot; la lame d'acier de la guimbarde est mise en vibration entre les dents; on entend une parole parfaitement intelligible, quoique " monocorde ". L'expérience montre lors des analyses que le spectre de raies reste fixe : ce n'est donc pas le spectre de raies qui véhicule l'information sémantique.

Mais on peut également fabriquer une parole parfaitement intelligible de façon encore plus simple en supprimant le spectre de raies de la parole normale par simple déconnection des cordes vocales, la " soufflerie " et les modifications formelles de l'organe phonatoire étant conservées. C'est alors tout simplement la voix chuchotée. On vérifie facilement que celle-ci conserve une intelligibilité totale. Des expériences que nous avons faites à partir de cette constatation nous ont montré l'intérêt théorique extraordinaire de la voix chuchotée, dont nous avons dès lors entrepris une étude systématique. A l'aide de diverses opérations (filtrage en particulier) on peut encore supprimer une certaine quantité d'énergie inutile à l'intelligibilité de la voix chuchotée. Finalement, entre parole normale et parole chuchotée juste encore intelligible, on observe une chute d'au moins 60 dB. La quantité d'énergie représentant l'information sémantique dans la parole normale est donc absolument infime. On vérifie au sonagraphe qu'il ne reste sur le document que des traces; mais le point capital est que ces traces présentent tous les caractères d'une forme au sens de la Gestalttheorie.

V - LES MOTS SONT DES FORMES

La parole résulte de l'enchaînement de mots; les mots sont les unités phonétiques du langage. Nous avons donc commencé par établir des sonagrammes de mots en parole chuchotée filtrée. Nous avons alors

...../

observé que tout ce que la gestalttheorie avait établi pour les formes visuelles reste valable dans le domaine de la parole. La raison en est fort simple : nous avons depuis longtemps la conviction - et les physiologues possèdent de ce point de vue de nombreux indices - que la parole aboutit d'abord à la projection très matérielle d'un véritable " sonagramme " sur le cortex. Il n'y a donc aucune différence entre formes visuelles, acoustiques, tactiles, etc... puisque dans tous les cas nos sensations aboutissent à une projection corticale électrique en trois dimensions (intensité, composition " spectrale " et temps) qui, automatiquement quantifiée, est alors susceptible d'être traitée par l'organe central du cerveau sous cette forme. En d'autres termes, toutes nos sensations sont des formes, et les lois de la Gestalttheorie s'appliquent à tous nos sens. Parmi ces lois, on vérifie aisément les suivantes, en considération de ce qui précède :

a) une forme acoustique est transposable dans le domaine de la parole; la transposition peut se faire en fréquence, en intensité ou en durée ; la "forme " sonographique d'un mot (stéréotypée : telle que nous la conservons dans notre mémoire, et qui représente l'étalon de référence inscrit dans notre mémoire mémorisante lors de l'apprentissage - où nous apprenons précisément à extraire l'information sémantique pure), cette forme peut être placée plus haut dans l'échelle des fréquences ou plus bas (selon les dimensions des cavités); elle peut être allongée ou présenter des proportions différentes; mais elle reste toujours reconnaissable, dans la mesure où les marges de fluctuation ne dépassent pas les marges statistiques de dimensions et de mouvements de l'appareil phonatoire humain . Sinon, nous sommes déroutés (fig.2) et obligés d'apprendre un code un peu différent mais très similaire, comme nous l'avons vérifié.

En tous cas, les idées peuvent être véhiculées par n'importe quelle sensation, puisqu'on y dispose de trois dimensions variables pour chacune d'elles; seulement l'organe phonatoire et l'organe de l'ouïe sont beaucoup plus élaborés que l'organe du goût ou de la vue, par exemple : il n'est donc pas surprenant que l'homme ait instinctivement choisi ses organes " acoustiques " pour véhiculer des formes aussi compliquées que la pensée.

...../

En bref, les formes " pures " des mots telles que nous les obtenons par voix chuchotée-filtrée peuvent être plus grandes, plus petites, plus ou moins anamorphosées, placées plus ou moins haut, etc.: elles restent reconnaissables dans de larges limites.

b) Une forme n'est pas une somme d'éléments (les phonèmes) mais une fonction de plusieurs variables. Le phonème est bien l'"atome" de la parole, mais il change de forme selon la situation et la fonction qu'il occupe dans le mot. On vérifie aisément sur le sonagramme (fig.3) que le mot " oui " n'est pas la succession des sons "ou" et "i", mais une forme " linéaire " originale (courbes allant de ou vers i). Or les essais de " parole pure " que nous avons faits ont montré avec évidence qu'un mot est une forme globale, mais que l'on peut décomposer précisément en formes de liaison raccordées, qui sont les véritables unités élémentaires de la parole. Nous appelons ces unités " éléments phonétiques "; ils traduisent en fait le mouvement global du système phonatoire entre deux positions limites (qui sont les phonèmes).

En bref, un mot, considéré comme forme globale, n'est pas la somme des phonèmes, mais l'association, le raccordement des éléments phonétiques.

Ainsi le mot parole n'est plus découpé en 6 phonèmes (p, a, r, o, l, e) mais en cinq éléments phonétiques : pa-ar-ro-ol-le dont le raccordement est automatique si nous utilisons des sonagrammes d'éléments phonétiques stéréotypés. Nous avons donc été conduits à établir le dictionnaire de ceux-ci.

VI - LE DICTIONNAIRE DES ELEMENTS PHONETIQUES

Comme base de départ, nous avons pris la liste des phonèmes de la sténographie DUPLOYE, dont l'expérience montre qu'elle permet, avec les signaux graphiques correspondants, d'écrire tous les mots de la langue française. Ces phonèmes sont au nombre de 31.

...../

En les associant deux à deux, on obtient l'ensemble des éléments phonétiques de la langue française. Il en existe donc 961.

On réalise une matrice d'ordre 31 à deux entrées, avec les phonèmes en ligne et en colonne.

Pour des raisons pratiques, nous avons été amenés à nous intéresser aux taux d'occurrence des éléments phonétiques; certains d'entre eux sont très fréquents, d'autres très rares. Nous avons donc pris des textes (articles de journal, poésie, etc...; nous avons découpé ces textes en éléments phonétiques et établi leurs taux respectifs d'apparition. Par la suite J.S. LIENARD a systématisé les opérations en établissant un programme de dépouillement sur machine numérique (CAB 500). L'expérience a montré que les taux d'occurrence variaient selon les textes et les auteurs; il est différent pour un article scientifique et pour une poésie, mais dans l'ensemble on peut dégager une liste statistique moyenne des éléments phonétiques qui suit à peu près la loi de Zipf (logarithmique). En millièmes, on trouve les fréquences d'apparition suivantes, par ordre décroissant :

de (17) ; ar (15) ; sieu (14) ; la (13) ; as (12)
le (11) ; il (11) ; li (10) ; ion (10) , etc....

Cette classification nous a permis de comprendre l'allure phonétique de la langue parlée française; nous avons également fait des sondages dans d'autres langues.

L'établissement des sonagrammes de ces éléments phonétiques étant un travail très long, nous avons ainsi pu travailler avec les plus importants, remettant à plus tard le travail complet sur les éléments phonétiques rares.

Pour extraire les formes stéréotypées nécessaires, nous avons enregistré successivement en voix chuchotée-filtrée les sonagrammes des éléments phonétiques en prenant nos trois voix (une féminine).

..../

Des sonagrammes obtenus, nous avons extrait, par confrontation, une forme moyenne, schématisée à l'aide d'un pinceau et d'encre de chine. Nous avons ensuite quantifié cette forme en la redessinant sur un carton quadrillé par petits carrés juxtaposés. L'intérêt de cette opération apparaîtra plus loin.

Finalement, nous avons en mains la partie la plus importante du dictionnaire des formes stéréotypées des éléments phonétiques les plus fréquents. Ce dictionnaire nous a servi, d'abord à vérifier nos hypothèses relativement à l'intelligibilité de la parole pure dessinée à partir des éléments phonétiques; ensuite à passer au stade de la synthèse de la parole, qui découle directement de ces hypothèses.

VII - LA SYNTHÈSE DE LA PAROLE PURE

Pour vérifier nos hypothèses et passer à la synthèse de la parole pure, nous avons été conduits à imaginer un appareillage de synthèse approprié, qui a été construit puis mis au point et utilisé : l'ICOPHONE.

C'est exactement l'inverse du sonographe. La parole pure est dessinée sur une bande de papier que l'on dispose sur un cylindre tournant. Une cellule photo-électrique explore la bande du bas en haut par spires successives serrées. Lorsque la cellule passe devant un point noir, elle déclenche un signal électrique dont la fréquence correspond à celle du point considéré. Ce signal (quasi-sinusoïdal) est ensuite enregistré sur un disque magnétique solidaire du cylindre; on se construit ainsi, par " re-recording " l'ensemble de la " forme " graphique. Le signal global que l'on peut écouter en fin d'opération correspond à la parole pure.

Nous ne sommes pas habitués à cette parole qui possède un caractère particulier; mais c'est de la parole chuchotée et les mots sont parfaitement intelligibles moyennant une certaine habitude. Avec des retouches, il sera donc possible, à l'aide du dictionnaire des

formes stéréotypées, de composer des mots sans passer par l'analyse préalable de ceux-ci : Il s'agit en fait d'une espèce de sténographie acoustique. En tout cas, les expériences sont concluantes et les conséquences que cela peut entraîner dans le domaine de la parole en général nous semblent de la plus haute importance. En particulier :

- il devient possible de faire la synthèse de la parole avec des moyens économiques (tel notre appareil actuel), ce qui facilitera les études de phonétique, et de parole en général.
- la synthèse d'une parole intelligible avec machines à calculer ayant une sortie " son " est immédiate à partir des formes stéréotypées quantifiées des éléments phonétiques.
- l'étude des défauts de la parole, la rééducation peut grandement bénéficier de la notion de " parole pure ".
- les télécommunications peuvent y trouver un intérêt d'économie dans les moyens qu'elles utilisent.
- on peut entreprendre sur des bases solides des études de linguistique variées.
- d'autres horizons peuvent s'ouvrir dans le domaine de la commande vocale de machines (la reconnaissance des formes simples constituant les éléments phonétiques est facilitée).

VIII - CONCLUSIONS

Notre travail sur la parole n'est pas terminé; l'ICOPHONE tel que nous l'avons réalisé est un appareil de laboratoire un peu trop rudimentaire; difficile à régler, il demande une grande patience et beaucoup de temps est perdu en manipulations délicates; c'est pourquoi nous n'avons pas réalisé le dictionnaire des éléments phonétiques dans son intégralité. D'autre part, il est évident que la " forme "

des éléments phonétiques, parce qu'elle est épurée au maximum et dépouillée de sa redondance devient extrêmement fragile et est facilement détruite par les parasites de notre appareillage actuel. Mais nous avons en cours de construction un appareil plus maniable, plus fiable et plus rapide qui nous permettra de développer toutes les études que nous avons entrevues, et en particulier de terminer le dictionnaire des formes stéréotypées. Nous espérons ainsi pouvoir apporter une contribution notable aux problèmes si importants de la parole auquel travaillent avec acharnement de nombreux laboratoires dans le monde entier.

E. LEIPP - Paris le 25 Mai 1966

Nous désirons exprimer ici notre gratitude à M. le Professeur SIESTRUNCK, Directeur du Laboratoire de Mécanique Physique de la Faculté des Sciences de Paris qui nous a fourni les moyens sans lesquels ce travail n'aurait pu être réalisé et nous a constamment encouragés aux inévitables moments difficiles qui se présentent lors de toute recherche.

M. E. LEIPP

Melle CASTELLENGO

M. J.S. LIENARD