

Groupement des Acousticiens de Langue Française

Texte des tables rondes
des

9^{èmes} JOURNÉES D'ÉTUDES SUR LA PAROLE

du

Groupe de la

" COMMUNICATION PARLÉE "

VOL. 2

LANNION

31 mai - 2 juin 1978

enet

BP 40
22301 LANNION Cedex
France

Edition : Décembre 1978

Groupement des Acousticiens de Langue Française

Texte des tables rondes
des

9^{èmes} JOURNEES D'ETUDES SUR LA PAROLE
du
Groupe de la
" COMMUNICATION PARLEE "

LANNION

31 mai - 2 juin 1978

enet

BP 40
22301 LANNION Cedex
France



Edition : Décembre 1978

Institut de Phonétique
Inventaire n° 2085
Cote n° A/SEP 9/2

Cette brochure contient les textes rédigés par deux rapporteurs des tables rondes qui se sont tenues au cours des 9èmes Journées d'Etudes sur la Parole du 31 Mai au 2 Juin 1978 à Lannion.

La première table ronde avait pour sujet :

"Stimulus de laboratoire et parole naturelle"

(Peut-on transposer à la parole naturelle les résultats d'expériences réalisées à partir de stimuli artificiels ?).

Les modérateurs étaient Elisabeth LHOTE (Institut de Phonétique de Besançon) et Max WAJSKOP (Institut de Phonétique de Bruxelles). C'est Max WAJSKOP qui en fait ici le compte rendu.

La seconde table ronde avait pour sujet :

"Evaluation de la qualité et de l'intelligibilité de la parole synthétique".

Les modérateurs étaient Jacques GENIN (Département E.T.A. - C.N.E.T.-Lannion), Joël LE ROUX (E.N.S.T. Paris) et Daniel TEIL (L.I.M.S.I., Orsay). C'est Jacques GENIN qui en fait ici le compte rendu.

Les textes des rapporteurs des trois thèmes de ces Journées d'Etudes sont publiés dans la Revue d'Acoustique n° 47 de Janvier 1979.

R. DESCOUT.



Table Ronde

Stimulus de laboratoire et parole naturelle

Peut-on transposer à la parole naturelle les résultats d'expériences réalisées à partir de stimuli artificiels?

Max WAJSKOP et Elisabeth LHOTE, animateurs de cette Table Ronde, exposèrent d'abord leur point de vue sur ce problème avant d'ouvrir la discussion générale. Max Wajskop d'est chargé de résumer les introductions et le débat.

Max WAJSKOP (Phonétique - Bruxelles)

Il n'est pas inutile de rappeler que le thème de cette Table Rondé est né d'un souhait émis en conclusion de la journée de formation sur la méthodologie des tests de perception. Si la question posée n'est pas triviale, certains de ses aspects sont cependant situés au centre d'un faux débat. Elle résulte en effet avant tout d'un sentiment de frustration qu'il faut dissiper et dont les sources peuvent être diverses. On peut dénombrer trois causes à ce malaise.

- 1°. Les phonéticiens non habitués à la tradition de l'expérimentation perceptive sont quelque peu agacés de constater que l'essentiel de la recherche de ces 20 dernières années se fonde sur des résultats obtenus à l'aide de séquences composées de quelques segments phonétiques.
- 2°. D'autres (une minorité), tout en reconnaissant la validité expérimentale de ces travaux, mettent en cause l'utilisation de stimuli synthétiques dont la composition et la structure épurée sont loin de représenter à leurs yeux la richesse et la complexité de la parole naturelle même lorsqu'elle est réduite à une suite de quelques phonèmes.
- 3°. D'autres chercheurs, enfin, estiment que les résultats obtenus sur des séquences aussi limitées ne peuvent être extrapolés à la parole continue qui présente des patrons ou qui contient des indices absents ou insuffisamment représentés dans les séquences de laboratoire. Avec cette troisième catégorie de critiques, on rejoint en fait le débat

qui oppose les tenants d'une approche globale de la perception de la parole et ceux qui estiment que la perception de la parole, du moins jusqu'à un certain niveau, ne se différencie guère de la perception psycho-acoustique.

Je me propose dans cette ouverture de m'en tenir à trois thèmes :

- 1°. Les raisons qui justifient l'emploi de stimuli artificiels de faible étendue.
- 2°. Les limites que l'emploi de tels stimuli impose aux implications ou aux déductions que l'on peut extraire des résultats expérimentaux. En particulier, on touchera ici au statut de la parole synthétique dans la recherche phonétique.
- 3°. Quels sont les problèmes que l'on peut rencontrer en passant du stimulus artificiel à la parole continue naturelle.

+
+ +

Jusqu'à l'apparition des synthétiseurs, la phonétique a été essentiellement une discipline instrumentale et descriptive. La synthèse de la parole allait lui permettre de tester un certain nombre d'hypothèses et surtout de commencer à établir des liens, sinon des relations causales, entre les différents niveaux du circuit de la communication parlée. Au début des années cinquante, cette orientation s'inscrit à côté du développement des notions de trait distinctif et d'indice.

Les questions que l'on se pose à cette époque sont déjà typiques de celles qui sont reprises aujourd'hui. Quels sont les indices responsables de l'accent anglais? Quels sont les formants responsables de l'identification vocalique? Quel est l'indice responsable de l'identification du point d'articulation des consonnes occlusives? Quel est l'indice qui détermine l'opposition de voisement des occlusives? etc...

Pour répondre à ces questions, on pouvait choisir trois voies différentes :

- a) la description analytique rigoureuse de corpus phonétiques de grand volume et l'extraction de schèmes invariants.
- b) l'utilisation de la synthèse
- c) une démarche complémentaire utilisant en parallèle la description et l'investigation perceptive.

Durant près de 15 ans, c'est presque exclusivement la deuxième démarche qui a été utilisée.

Parmi les diverses raisons qui peuvent expliquer cet engouement, il en est une qui a particulièrement séduit les chercheurs. La technique qui consiste à utiliser de la parole synthétique satisfait aux critères de la méthode scientifique. Elle aboutit en effet à des observations contrôlées et vérifiables. Dès lors, si on reprend la définition de Piaget pour qui le fait scientifique se définit comme la réponse vérifiée à un problème, c'est-à-dire une mise en rapport de relations répétables au sein d'une démarche formelle, on aboutit à une démarche qui présente apparemment tous les signes voulus de rigueur.

L'utilisation de stimuli synthétiques dans des expériences psychologiques obéissant elles-mêmes à des consignes éprouvées permettait de répondre aux caractéristiques que je viens d'énoncer. De plus, la parole synthétique permet de représenter au plan du réel cette démarche formelle. Elle sert donc à construire un modèle qui pourra rendre compte du système de relations que le chercheur établit.

Tout ceci paraît donc justifier l'exploitation de stimuli artificiels. Quant à l'emploi de courtes séquences durant ces recherches, il s'explique par des raisons de caractère pragmatique :

- a) la faiblesse technique des premiers synthétiseurs et le caractère peu naturel de leur production
- b) la nécessité d'établir des références calibrées, des normes avant de s'aventurer dans des champs plus vastes.

Ce dernier point mérite peut-être un développement car il n'est pas certain qu'il rencontre l'unanimité.

Tant au niveau de la description phonétique qu'à celui de l'expérimentation perceptive, j'ai toujours défendu l'idée qu'il fallait établir des cadres de référence. Ceci suppose et réclame, même dans l'analyse de la parole naturelle continue, un certain degré d'artificialité qui permette un contrôle minimal de paramètres.

Cette démarche plus longue et qui réclame beaucoup de précautions, notamment dans le choix et le nombre de sujets, dans les procédures techniques, a cependant l'avantage d'établir les fondations d'un édifice. Il en va de même pour les recherches de caractère psycho-acoustique ou psycho-phonétique. Après tout, c'est ainsi que la psycho-acoustique s'est bâtie et si aujourd'hui on est obligé de réviser un certain nombre de soi-disant découvertes en perception de la parole, c'est

parce que des chercheurs formés aux techniques de la psycho-acoustique ont réappliqué avec vigueur ces techniques à des thèmes que l'on croyait solidement établis. (cf. travaux récents de J.D. MILLER au Central Institute of the Deaf à Saint Louis).

En résumé, il me paraît que l'utilisation de stimuli artificiels de faible étendue se justifie aisément.

Ces stimuli permettent de modéliser le réel; ils peuvent servir de cadres de référence à des examens plus vastes; ils ont abouti à éclairer nombre de questions en perception de la parole. Aussi bien pour ceux qui estiment que la perception se présente sous la forme d'une démarche, soit ascendante, soit descendante, l'élucidation des questions qui se posent aux premiers niveaux de la réception du signal sonore est absolument nécessaire.

+
+ +

Mais de tels stimuli n'imposent-ils pas de limites sévères aux conclusions que l'on déduira des résultats expérimentaux?

Qu'appelle-t-on un stimulus artificiel?

Tout signal naturel auquel on a imposé une modification (amplitude - fréquence - temps) et tout signal fabriqué par l'homme. Si cette définition est acceptable et acceptée, on conviendra que pratiquement tous les signaux utilisés au laboratoire sont des stimuli artificiels. Les différences qui peuvent exister entre eux ne sont pas de nature qualitative. On doit parler plutôt de degrés d'artificialité.

Exemples: 1-sujet humain: fo constant, amplitude constante, rythme d'élocution

2-production humaine modifiée: excision, intensité, tempo, fondamentale, vocoder

3-production synthétique

C'est évidemment la parole synthétique qui est le plus souvent visée. En 1962, Fant disait déjà qu'elle avait soulevé plus de problèmes qu'elle n'en avait résolus et que les conclusions qu'on pouvait en tirer à l'époque ne devaient s'appliquer qu'à la perception de la parole synthétique et non à la perception de la parole naturelle.

La parole synthétique nous fournit un modèle au sens premier du terme: c'est un instrument de contrôle des différents paramètres. A ce titre, c'est un admirable instrument de travail. Mais elle peut présenter des dangers. On peut en effet mettre en évidence n'importe quoi,

n'importe quel indice à partir du moment où dans une situation épurée, dénuée de toute redondance (c'est-à-dire tous les autres indices étant neutralisés) on ne fait plus varier qu'un seul indice à la fois.

Les résultats que l'on obtient ainsi sont des plus utiles sur le plan psycho-physique mais il convient de vérifier avec soin dans quelle mesure ils peuvent être étendus à la parole naturelle.

Elisabeth LHOTE (Phonétique - Besançon)

Après cette défense et illustration des stimuli artificiels, je voudrais défendre l'emploi de la parole continue naturelle. Ma position sera celle d'une remise en cause : quelle sera l'attitude d'un phonéticien face aux phénomènes sonores qui l'entourent, aux diverses "paroles" qui l'environnent? C'est donc à une suite d'interrogations que je me livrerai.

Qu'est-ce que la parole naturelle? Cette notion n'évoque pas la même chose pour le phonéticien, l'ingénieur et le médecin. Pour l'ingénieur, "parole naturelle" doit signifier probablement l'insertion de l'humain, de l'individuel dans le flux de la parole. Par ailleurs, le stimulus artificiel ne se définit pas seulement par sa limitation dans le temps. De plus, tout le monde sait que nous ne réagissons pas de la même façon vis-à-vis du mot isolé ou de ce même mot inséré dans la phrase. En intonation par exemple, c'est l'ensemble de la séquence qui doit être considéré. Je me réfère également à ce que Di Cristo a appelé "l'effacement perceptuel", c'est-à-dire le rôle que jouent de petites unités dans l'intégration perceptive d'unités de grande dimension. Ce problème ne peut être abordé à l'aide de stimuli artificiels réduits. Enfin, je crains qu'en travaillant uniquement à l'aide de stimuli synthétiques, c'est-à-dire à l'aide du produit fini observable, on néglige tout l'aspect génétique du circuit de la communication.

DISCUSSION

JUBAN (U.E.R. Linguistique - Brest)

Insiste sur la nécessité d'établir une distinction entre "parole" et tout "bruit" produit par l'homme et/ou par la machine. Le travail qui opère uniquement sur l'observable physique ou au mieux sur l'observable psychophysique efface une grande partie de l'élaboration du comportement verbal. JUBAN récuse une investigation dont les

modalités dérivent des techniques expérimentales de l'ingénieur des télécommunications. Il en est résulté une confusion permanente entre "perception" et "reconnaissance" par l'homme et par la machine. Il faut rompre avec l'expérimentation instrumentale et ses contraintes technologiques pour aborder une expérimentation clinique. Les troubles pathologiques classés globalement sous le terme d'agnosies : amusie, surdit  musicale, aphasie de conduction, changement de voix du schizophr ne constituent un terrain d'excellence pour le chercheur car ces troubles d montrent souvent que leur apparition est li e   l'insertion du sujet dans des contextes non naturels. Ces pathologies aux aspects tr s diff renci s r clament une r incorporation de l'humain dans la recherche car percevoir le langage ce n'est pas percevoir le train qui passe.

de BOYSSON (Paris)

Met l'accent sur le statut de l'objet trait  et celui de la r ponse en rappelant que le stimulus artificiel est un essai de reproduction de la parole naturelle mais que nous ne disposons pas d'une th orie de la ressemblance qui nous permette de poser correctement le statut de la parole artificielle. De mani re analogue, on n glige le fait que la r ponse du sujet est toujours r ponse  labor e. Le crit re ou le rep re sur lequel se fonde le sujet vis- -vis du stimulus artificiel n'est g n ralement pas le m me que celui sur lequel il s'appuie face   la parole naturelle. Le sujet ne hi rarchise pas de la m me mani re vis- -vis de ces deux cat gories.

SORIN (C.N.E.T.)

Rel ve   propos de la notion d'artificiel, ce qui distingue la production de la perception et rappelle que m me si le stimulus est de la parole continue et naturelle, la r ception en laboratoire est une condition d'audition non naturelle, d'o  l'apparition de modes diff rents de r ponse, d'ancrages, de mise en place de proc dures de r ponse plus ou moins naturelles. Il lui para t plus important de se pencher sur la mani re dont on pourra prendre en compte les r troactions des niveaux sup rieurs sur le niveau p riph rique dont le r le semble s'affirmer. Mais cette r troaction est difficilement scrutable si l'on se contente d'utiliser des stimuli peut- tre calibr s mais r duits.

DELGUTTE (M.I.T.)

Rappelle d'abord que la parole synthétique de haute fidélité, personnalisée, existe désormais ce qui place l'expérimentateur dans de meilleures conditions de travail. En ce qui concerne le problème méthodologique posé aujourd'hui et les divergences dans les points de vue évoqués, Delgutte souligne que des techniques d'investigation appartenant à des disciplines différentes sont de plus en plus empruntées par les unes et par les autres. Des résultats identiques obtenus par des techniques différentes forment dès lors des noeuds de convergences qui sont assurés d'une vérité mieux fondée.

LHOTE (Phonétique - Besançon)

La parole synthétique "personnalisée" ne remplace pas le dialogue avec le locuteur véritable et rien ne permet d'assurer que la réponse du sujet, dans le premier cas, soit identique à celle qu'il fournirait dans le second.

WAJSKOP (Phonétique - Bruxelles)

La position adoptée par JUBAN n'est pas antagoniste de celle adoptée par l'expérimentation instrumentale ou analytique, elle en est complémentaire et traitée dans d'autres laboratoires. Il en va de même pour la thèse de LHOTE. S'il y a un statut du sujet et un statut de la réponse, il y a aussi un statut de la question posée. Face à la parole synthétique, le sujet sait qu'il n'est pas dans une situation de dialogue. C'est évidemment sur les noeuds de convergence cités par Delgutte que la recherche peut se redéployer grâce à des questions autrement formulées : la progression en spirale de la recherche phonétique depuis 1945 avec intégration successive des résultats antérieurs est un exemple de ce genre de développement analogue à celui d'autres sciences. Par contre, nous ne disposons pas toujours, comme les physiciens, de théories dénuées de troubles terminologiques.

CARTON (Phonétique - Nancy)

Met l'accent sur la démarche dialectique entre stimulus bref et stimulus long (parole continue) au même titre qu'il y a complémentarité entre la linguistique de la phrase et la linguistique textuelle.

Cette dernière faisant surgir des questions non posables au travers de la phrase isolée.

ROSSI (Phonétique - Aix)

Définit le stimulus naturel comme de la parole spontanée se déroulant dans une activité de langage. Si les chercheurs avaient tenté de travailler immédiatement à ce niveau, on en serait encore au Moyen Age. Les chercheurs sont conscients, et des limites d'emploi, et des limites de généralisation lorsqu'ils utilisent des stimuli artificiels. Rossi énumère ensuite les difficultés d'emploi de la parole artificielle en laboratoire. Il montre que le manque d'une théorie de la ressemblance peut être contourné à l'aide de tests préalables de compréhension et/ou de naturel. En ce qui concerne le statut de la réponse, il est précisément intéressant de voir comment la consigne et la méthode (méthode des limites, méthode des stimuli constants, méthode ABX) influent sur le comportement du sujet. Troisième difficulté : la création éventuelle d'un cadre de référence interne, d'un apprentissage du sujet. Autre difficulté, la plus grave peut-être, la recreation d'une situation langagière induisant chez le sujet des discriminations non contrôlées par l'expérimentateur.

Il est évident, selon ROSSI, que les démarches doivent être complémentaires mais que nos connaissances en perception de la parole ne seraient pas aussi avancées si l'on n'avait pas disposé de résultats obtenus sur des stimuli synthétiques. Ce qui compte aujourd'hui c'est de déceler les résultats convergents et de constater, par exemple, que les seuils différentiels de durée sont plus bas encore en parole continue que ceux que l'on avait obtenus en laboratoire.

Après ces interventions, s'est déroulée une discussion générale qui a porté essentiellement sur les avantages et désavantages des stimuli synthétiques, l'adéquation entre les caractéristiques de la production vocale humaine et les copies artificielles, le problème des "monstres" acoustiques. Le débat a ensuite rebondi sur les thèmes soulevés par JUBAN.

Faute de temps, le 3e sous-thème "Peut-on transposer à la perception de la parole naturelle les résultats obtenus en laboratoire? Si oui, comment?" n'a pu être abordé.

9èmes JEP - 1978

LANNIONTABLE RONDE

<p>"EVALUATION DE LA QUALITE ET DE L'INTELLIGIBILITE DE LA PAROLE SYNTHETIQUE".</p>

INTRODUCTION

L'exposé qui suit n'est pas une transcription fidèle des débats tels qu'ils se sont déroulés. L'ordre chronologique même dans lequel les questions ont été abordées n'est pas respecté. Je me suis efforcé de réordonner les idées émises et discutées afin d'en obtenir la présentation la plus objective.

TERMINOLOGIE

Le mot "intelligibilité" se suffit à lui-même. Par contre la notion de "qualité de la parole synthétique" est moins facile à saisir. On s'interroge également sur l'absence, dans l'énoncé du thème, du mot "naturel".

"Qualité" apparaît comme notion plus générale, regroupant trois concepts différents : "intelligibilité", "agrément", "naturel".

"Agrément" est la réponse à la question : cette voix est-elle ou non agréable ?

"Naturel" est la réponse à la question : cette voix semble-t-elle avoir été prononcée par un être humain ?

INTELLIGIBILITE

Les laboratoires représentés ici qui ont conduit des tests d'intelligibilité sur de la parole synthétique sont essentiellement le LIMSI (tests sur mots aléatoires sans signification) et le CNET (test sur 100 noms propres, tests d'énoncés de lignes d'annuaire, test prévus sur le système de dénumérotation à Saint Briec).

.../...

Le but recherché est soit de s'assurer d'une intelligibilité suffisante, soit d'améliorer le système dans le sens d'une meilleure intelligibilité. Jusqu'à présent aucune étude n'a été conduite visant à comparer des systèmes différents.

On demande à ces études de guider l'amélioration des synthétiseurs, mais, ici existent aussi les méthodes utilisées en transmission de la parole. On leur demande également d'améliorer la méthode de synthèse (dictionnaire ou règles, gestion de la prosodie, choix de la voix analysée pour servir de référence).

QUALITES D'AGREMENT ET DE NATUREL

Pour certains (LIMSI), la seule intelligibilité importe, agrément et naturel ne sont que des arguments commerciaux. D'autres au contraire mettent l'effort principal sur l'agrément et le naturel. La réponse dépend en fait du but que l'on se fixe. Pour une application visant à attirer l'attention, le naturel de la voix risque d'être nuisible ; dans le cas de l'automatisation d'un service existant il est indispensable à la bonne "acceptabilité" de ce service ; parfois même la fonction intelligibilité devient secondaire (annonces dans les aéroports !).

METHODOLOGIE

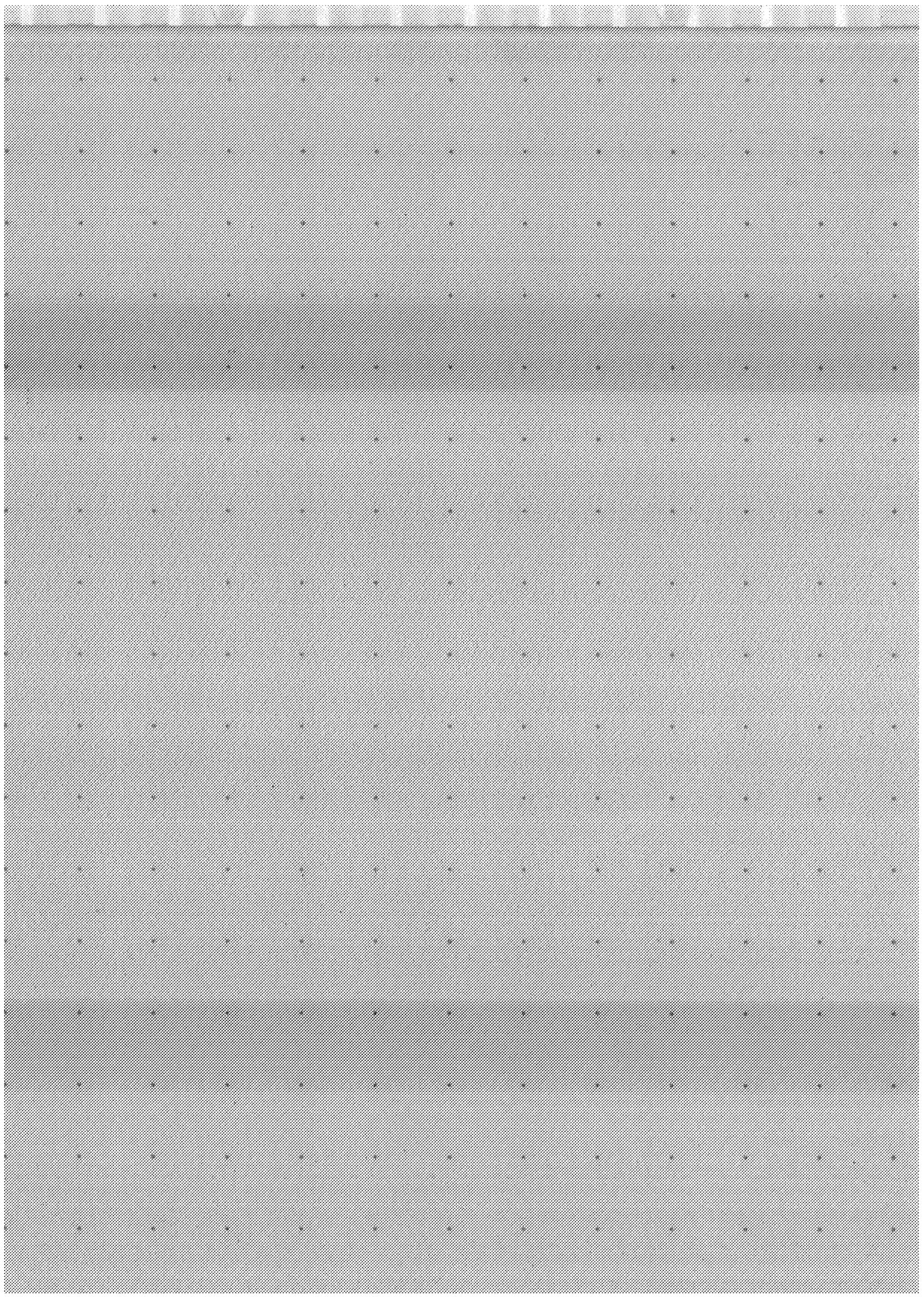
Au delà des méthodes initiales de test de mots présentés isolément, on constate que la séparation entre intelligibilité, agrément et naturel devient plus diffuse dans les procédés utilisés.

On trouve deux options dominantes.

D'une part, et on arrive ici à la notion "d'acceptabilité" de service, on cherche à évaluer un système opérationnel. On n'a pas alors le choix des signaux de parole à présenter. Se pose par contre le problème de l'obtention des résultats de tests. Comment interroger le public, quelles questions poser qui permettront d'obtenir, au delà d'un taux d'acceptabilité ou de rejet, des informations plus techniques pour corriger le système.

D'autre part, dans le cadre de mesures de laboratoire on cherche à ramener des mesures d'agrément à des mesures d'intelligibilité en masquant le signal de parole synthétique par du bruit. Cette méthode doit être utilisée avec précaution, elle fait intervenir une notion nouvelle, peut-être totalement différente de celles évoquées jusqu'ici : la résistance de la parole à la dégradation. Les paroles synthétiques, si elles sont mieux découpées et mieux articulées risquent de paraître de meilleure qualité que les paroles naturelles !.

.../...



CONCLUSION

Il est évident que cette table ronde est loin d'avoir épuisé le sujet, même d'en avoir totalement fait le tour. Beaucoup de travail est encore à fournir sur les mesures d'intelligibilité, plus encore sur les mesures de qualité avant que l'on puisse se faire une idée claire du problème.

J. GENIN