



NUMERO DE PUBLICATION : 1002379A4

NUMERO DE DEPOT : 8800993

MINISTERE DES AFFAIRES ECONOMIQUES

Classif. Internat.: G10L

Date de délivrance : 22 Janvier 1991

Le Ministre des Affaires Economiques,

Vu la loi du 28 Mars 1984 sur les brevets d' invention, notamment l' article 22;

Vu l' arrêté royal du 2 Décembre 1986 relatif à la demande, à la délivrance et au maintien en vigueur des brevets d' invention, notamment l' article 28;

Vu le procès verbal dressé le 31 Aout 1988 à 24h00
à l' Office de la Propriété Industrielle

ARRETE:

ARTICLE 1.- Il est délivré à : DE HALLEUX Benoit;DUTILLEUL Eric;JOFFROY Jean-Marc;
avenue Chevalier Jehan 82, 1300 WAVRE(BELGIQUE);rue du Calvaire 90, 6061
MONTIGNIES-SUR-SAMBRE (BELGIQUE);rue Vital Casse 18, 1490 COURT-SAINT-ETIENNE
(BELGIQUE)

représenté(e)(s) par : VAN MALDEREN Michel, OFFICE VAN MALDEREN, Avenue
J.-S. Bach, 22 bte 43 - B-1080 BRUXELLES.

un brevet d' invention d' une durée de 20 ans, sous réserve du paiement des taxes
annuelles, pour : PROCEDE DE RECONNAISSANCE ET EN PARTICULIER D'ECRITURE DE LA PAROLE.

ARTICLE 2.- Ce brevet est délivré sans examen préalable de la brevetabilité
de l' invention, sans garantie du mérite de l' invention ou de l' exactitude de
la description de celle-ci et aux risques et périls du(des) demandeur(s).

Bruxelles, le 22 Janvier 1991
PAR DELEGATION SPECIALE :


WUYTS L
Directeur.

PROCEDE DE RECONNAISSANCE ET EN PARTICULIER D'ECRITURE DE LA PAROLE

Objet de l'invention

La présente invention vise à réaliser un procédé qui
5 permet la reconnaissance de la parole et en particulier l'écriture de
la parole utilisant un alphabet phonétique naturel. Elle concerne
également les moyens et dispositifs permettant la mise en oeuvre de
ce procédé.

Applications de l'invention

10 Les applications de la reconnaissance de la parole sont
de deux types : le dialogue oral homme-machine et l'aide aux
malentendants.

Le présent mémoire descriptif qui est basé sur des
expériences réelles d'écriture de la parole destinée à des
15 malentendants décrira essentiellement à titre d'illustration cet
aspect particulier. Il est bien entendu cependant qu'il n'y est pas
limité.

Résumé de l'état de la technique

Précisons d'emblée qu'à l'heure actuelle, aucun système
20 n'est à même d'égaler l'homme dans la reconnaissance vocale.

A titre d'exemple de ce que l'on peut faire aujourd'hui
dans le domaine du dialogue homme-machine, on citera des
applications développées dans le cadre de ARPA Speech
Understanding System (applications militaires USA). Ces techniques
25 permettent un dialogue sommaire, basé sur un vocabulaire limité
(1000 mots) et sur une syntaxe simple et immuable.

Malgré les développements récents de l'informatique et
de l'électronique, en particulier en ce qui concerne le traitement du
signal et l'intelligence artificielle, ces systèmes restent d'un coût
30 prohibitif et on est encore fort loin d'un modèle de reconnaissance de
la parole aussi performant que l'être humain.

Pour ce qui concerne l'aide aux malentendants,
essentiellement deux moyens sont envisagés - outre l'implantation
de prothèses chirurgicales - pour aider les personnes sourdes.

35 Le premier s'appuie sur l'exploitation du sens tactile et
se concrétise par exemple par l'élaboration d'un bracelet vibreur qui
donne des informations sommaires telles que sonnerie de portes,
appel sur téléphone graphique (Minitel), ...

Le deuxième moyen plus développé, vise à supplanter le canal auditif devenu inactif par le canal visuel. Il regroupe ainsi toutes les techniques destinées à la transposition du son en images comme la lecture sur les lèvres, le langage gestuel, le téléphone graphique (ou Minitel), les systèmes d'éducation à la parole, ...

La lecture sur les lèvres est une première façon dynamique de comprendre le message oral mais présente malgré tout des ambiguïtés gênantes. En effet, il existe dans la parole beaucoup de sosies labiaux ; ainsi, des sons comme "p" et "b" ou "d" et "t" ne peuvent pas être distingués par la simple forme des lèvres.

Une solution proposée par le docteur Carnet (Gallaudat Collège, Washington 1967) pour lever ces indéterminations fut d'ajouter un code phonétique complémentaire mis en évidence par la forme et la position de la main du locuteur autour du visage. L'inconvénient de cette méthode appelée "cued speech" est que l'apprentissage du code tant par le locuteur que par le lecteur est nécessaire pour une compréhension mutuelle. Un autre inconvénient de la lecture sur les lèvres est qu'il faut voir le locuteur de face.

Le langage gestuel et le téléphone graphique possèdent eux aussi ces inconvénients majeurs ; les deux personnes en dialogue doivent connaître le code de communication gestuel ou doivent être équipées toutes deux de l'appareillage ad-hoc dans le cas du Minitel.

Ces techniques de reconnaissance pour sourds - mise à part la lecture labiale qui reste toutefois fort hasardeuse - ont tendance à marginaliser leur groupe restreint d'utilisateurs (malentendants, famille, enseignants) vis-à-vis de personnes qui ne se sentent pas concernées d'une manière ou d'une autre par le problème des handicapés de l'ouïe et qui de ce fait ne feront aucun effort pour appréhender ces techniques.

Il n'existe pas véritablement à l'heure actuelle de traducteur efficace du son susceptible de délivrer au sourd une autonomie d'utilisation indispensable à son intégration dans le monde des malentendants.

Le premier essai du genre fut entamé par les laboratoires Bell pendant la deuxième guerre mondiale et se solda par la mise au point d'un appareil baptisé "sonagraphe" ("visible speech") qui transposait pour la première fois la parole en images appelées "sonagrammes".

Les sonagrammes sont des diagrammes à deux dimensions où l'abscisse est le temps et l'ordonnée représente les fréquences ; l'énergie sonore aux différentes fréquences étant matérialisée par une intensité de noirceur.

5 Depuis 40 ans, des expériences se multiplient un peu partout dans le monde pour tester les qualités de cette représentation mais le bilan s'avère en définitive plutôt décevant.

10 Il semble que la meilleure performance de lecture des sonagrammes appartient à Victor Zue, expert en reconnaissance de la parole au MIT (Boston). Celui-ci après une durée d'apprentissage de plus ou moins 2500 heures (quelques heures par jour pendant plusieurs années) parvient de la sorte à identifier 67 % des sons émis par un grand nombre de personnes et ce, en un temps 20 fois plus grand que le temps réel de production du message verbal.

15 Les mauvaises performances de lecture de sonagrammes résultent en grande partie d'une qualité médiocre voire inadaptée de la représentation utilisée ; l'emploi d'une intensité de gris approximant la quantité d'énergie en fréquence donne une image grossière de la réalité sonore et court-circuite ainsi une partie
20 importante de l'information phonétique. L'absence d'arrêt de l'évolution de l'image est la cause principale de la difficulté de lecture ; de plus, la redondance d'information ne permet pas de se concentrer sur la partie importante de la représentation visuelle.

But de l'invention

25 L'invention vise à fournir une représentation visuelle de la parole qui est de nature à surpasser largement le sonagramme tant par la plus grande richesse de l'information phonétique qu'elle présente que par l'aisance relative au niveau de son apprentissage.

Éléments caractéristiques de l'invention

30 L'invention repose sur un procédé de reconnaissance de la parole et en particulier d'écriture de la parole comportant une analyse fréquentielle des phonèmes émis par un locuteur, caractérisé en ce que les phonèmes subissent un traitement d'analyse fréquentielle spécifique qui pour chaque phonème exprimé génère une
35 image statique, apparaissant immobile (figée) pendant un temps suffisant pour permettre sa reconnaissance. De manière avantageuse, l'image statique représentant le phonème que nous appellerons le caractère-image doit apparaître sur l'écran d'une façon qui paraisse

instantanée et non progressive et rester immobile pendant le temps d'identification du phonème.

On s'est aperçu qu'un utilisateur peut après un délai d'apprentissage raisonnable reconnaître les différentes images
5 statiques ainsi associées à chaque phonème.

Selon une forme d'exécution préférée de la technique précitée, il est avantageux de prévoir un affichage simultané de plusieurs phonèmes successifs, ce qui permet à l'utilisateur de mieux reconnaître ceux-ci par association avec le ou les phonèmes
10 exprimés antérieurement à celui qui est lu et éventuellement le ou les phonèmes exprimés déjà après celui en cours de lecture.

Comme dans l'écriture classique, un nombre suffisant de phonèmes seront dans une application avantageuse de la technique visibles simultanément pour permettre la mémoire des éléments
15 verbaux précédents. Nous entendons par élément verbal, un phonème ou un groupe successif de phonèmes. Dans le cas où le mot est décomposé en éléments verbaux, le terme caractère-image se rapporte à l'image représentant l'élément verbal. Il est en effet souhaitable qu'au moins les caractères-images correspondant aux
20 éléments verbaux d'un mot, apparaissent au moment de l'interprétation du mot simultanément sur l'écran. En fait, il en va de même qu'avec l'écriture classique. Il suffit de songer aux conditions limites de lisibilité des sous-titres à la télévision et au cinéma. Il est souhaitable que les caractères-images soient immobiles pendant
25 leur apparition sur l'écran. Quant il s'agit d'écriture, le mot "caractère" lorsqu'il est utilisé sans qualificatif se rapporte à l'écriture phonétique et est une abréviation du mot "caractère-image".

Contrairement aux techniques connues, le caractère
30 statique de l'image associée à chaque phonème permet d'éviter la redondance de l'information, notamment pour les voyelles longues. La technique proposée aboutit à l'élaboration d'un alphabet naturel. A chaque élément verbal, phénomène essentiellement dynamique, est associée l'image statique (figée) qu'il génère grâce à un algorithme
35 particulier. Cet alphabet peut éventuellement être réversible, c'est-à-dire que les images peuvent à leur tour donner naissance aux types d'éléments verbaux qui les ont générés. On crée ainsi un alphabet "naturel" ou "physique" dans le sens d'une part ou un élément verbal

de par ce qu'il est, donne naissance à une image (le caractère-image) qui fait partie d'une famille d'images que l'on associe par éducation au type d'élément verbal en question et d'autre part que le caractère-image correspondant à un élément verbal peut générer par ce qu'il est, le type d'élément verbal en question.

Il s'agit d'un alphabet et la technique de reconnaissance de la parole par transposition en images est en fait une écriture à l'aide des images qui constituent les caractères. On conçoit donc que les caractères-images seront utilement groupés ensemble en suite successive dans un mot et que l'on séparera utilement par un espacement les caractères-images relatifs à des mots différents. Il s'agit des mots résultant de la parole et non des mots de l'écriture classique. Les mots de l'écriture phonétique la plupart du temps ne coïncide pas avec les mots de l'écriture classique. Souvent, le mot de l'écriture phonétique contient plusieurs mots de l'écriture classique. On entend par "mot" le mot de l'écriture phonétique à moins qu'il ne soit précisé qu'il s'agit du mot de l'écriture classique. On utilisera utilement et avantageusement une ponctuation particulière pour distinguer les espacements entre les mots, les espacements à virgule, point-virgule et point. L'évolution des caractéristiques adéquates du signal acoustique pendant la parole renseignera sur le caractère normal, interrogatif ou exclamatif d'une phrase, ce qui amènera naturellement à utiliser l'équivalent des points, points d'interrogation et points d'exclamation.

Une variante avantageuse de la technique comporte de plus l'affichage de la longueur de l'émission du phonème, de l'évolution dans le temps de son amplitude, de la régularité de l'émission acoustique du phonème dans le mot et dans la phrase et éventuellement des informations sur la fondamentale et les harmoniques.

Il est apparu que des résultats particulièrement intéressants sont obtenus par une technique comportant une opération d'analyse spécifique des phonèmes selon qu'ils appartiennent à un premier groupe constitué par les phonèmes autres que les consonnes fricatives ou les consonnes plosives, à un deuxième groupe constitué par les consonnes fricatives et un troisième groupe constitué par les consonnes plosives.

Il est apparu que chacun de ces groupes doit subir une analyse fréquentielle qui lui est propre, à savoir :

- une analyse fréquentielle du premier groupe au moins jusqu'à une valeur de l'ordre de 4 à 5 kHz sur une durée de l'ordre d'au moins
5 une quasi période,
- une analyse fréquentielle du deuxième groupe au moins jusqu'à une valeur de l'ordre de 12 kHz pendant une durée supérieure à 2 millisecondes et ne dépassant pas la durée des phonèmes,
- une analyse fréquentielle du troisième groupe au moins jusqu'à une
10 valeur de l'ordre de 5 à 10 kHz pendant une durée suffisante pour tenir compte de l'évolution temporelle des caractéristiques fréquentielles.

L'analyse fréquentielle pour les deux premiers groupes peut n'exploiter qu'une partie de l'information acoustique émise
15 pendant la production du phonème car cette information est redondante. L'analyse fréquentielle pour le troisième groupe exploite toute ou quasi toute l'information acoustique émise pendant la production de la plosive ou à tout le moins une partie notable de cette information.

20 Avantageusement des signaux analogiques captés par un micro subissent une digitalisation et sont soumis à l'aide d'un algorithme à une analyse fréquentielle propre au groupe afin d'obtenir une écriture de la parole basée sur l'analyse fréquentielle de celle-ci qui fournit un alphabet phonétique naturel.

25 Afin de faciliter la lecture de la dite écriture selon un alphabet phonétique naturel, l'invention comporte le fait que l'écriture correspondant à chacun des groupes précités subit une représentation particulière pour chaque groupe, notamment une représentation de couleur différente pour le groupe 1 de la couleur
30 retenue pour la représentation du groupe 2 tandis que les phonèmes du groupe 3 qui subissent une représentation du type sonagramme ou équivalente en ce sens que l'on exploite avantagement l'information acoustique présente dans quasi toute la durée du phonème, se représente avantagement dans une autre couleur que
35 les groupes 1 et 2.

On peut également discriminer les groupes en adaptant des conventions autres que les couleurs, par exemple des formes géométriques différentes pour chaque groupe. A titre d'exemple des

"caractères-images" de dimensions ou de formes différentes peuvent être utilisées.

5 Une façon avantageuse de présenter l'information de manière compacte est d'utiliser une représentation polaire des transformées fréquentielle ou équivalente.

10 On entend par traitement d'analyse fréquentielle ou analyse fréquentielle, un traitement mathématique de tout ou d'une partie du signal temporel lié au phonème, où le temps n'apparaît plus dans la résultante obtenue par la transformation et qui permet de mettre en évidence une information caractéristique spécifique à chaque phonème, qui sera utilisée pour reconnaître le phonème.

15 On entend par quasi-période, la période apparente du signal temporel quasi-périodique relatif aux phonèmes de la classe 1. La quasi-période est mesurée par le temps que dure le plus petit motif du signal temporel qui génère quasiment le signal temporel par juxtaposition successive. Le concept de quasi-période est illustré sur la figure 1.

20 D'autres détails et caractéristiques de l'invention apparaîtront à la lecture du mémoire descriptif qui suit, relatif à une forme d'exécution préférée de l'invention.

Mode d'exécution préféré de l'invention

La démarche suivie au cours de l'étude qui a conduit à la présente demande de brevet est essentiellement expérimentale.

25 Le critère utilisé pour apprécier les diverses représentations étudiées a été le pourcentage d'identification correcte des phonèmes par la personne qui lit l'écriture de la parole. Ce critère est sévère puisqu'il n'est généralement pas nécessaire d'identifier correctement la totalité des phonèmes pour comprendre le message. En effet, la signification, la cohérence et la structure syntaxique même du langage oral sont souvent déterminantes pour la bonne compréhension du message.

30 Le contenu phonétique du message est la seule information extérieure disponible à l'auditeur et c'est aussi la seule qui fait défaut au malentendant. C'est pourquoi un système d'aide visuelle ne doit pas "comprendre" le message mais simplement en assurer une représentation phonétique suffisante.

35 La représentation choisie est basée sur la notion de phonèmes et d'alphabet phonétique. En effet, l'information

phonétique, si elle varie continuellement au cours du temps, est la plupart du temps redondante localement de telle sorte que l'on peut décomposer le message oral en une succession de phonèmes et de parties de phonèmes. L'émission acoustique d'une partie d'un
5 phonème ou il y a redondance peut en effet caractériser entièrement le phonème.

La recherche s'est donc articulée autour de la représentation des différents phonèmes.

Ainsi, si chaque phonème peut être représenté
10 adéquatement, l'ensemble de ces représentations constitue alors un alphabet phonétique permettant, à l'instar de l'écriture de décrire le langage oral phonème par phonème.

L'invention repose donc essentiellement sur une représentation "naturelle" ou "physique" des phonèmes, plutôt que
15 sur une représentation symbolique de façon à garder la richesse et le caractère général de l'information naturelle que recèlent diverses nuances phonétiques (âge, sexe, prononciation, émotions, fatigue, ...) absentes dans une écriture où chaque phonème possède son symbole conventionnel stéréotypé (caractère-imprimé).

La classification des sons de la langue française adoptée
20 au cours de l'étude est fondamentalement liée aux modes de production des phonèmes. Cette classification regroupe en trois groupes l'ensemble des 36 phonèmes de la langue française. Le groupe 3 contient les plosives, le groupe 2 les fricatives et le
25 groupe 1 le reste des phonèmes soit les voyelles orales et nasales, les semi-voyelles, les consonnes liquides et nasales.

Cette classification rassemble les phonèmes dont les caractéristiques fréquentielles et/ou temporelles sont semblables. Elle est donc plus proche du but visé par l'invention que la
30 classification linguistique.

Dans chacun des groupes, certains phonèmes peuvent être regroupés de la façon suivante :

- a, â notés a (ex. : patte, pâte) ;
- e, ê, eu notés e (ex. : le, peur, mieux) ;
- 35 in, un notés ain (ex. : brin, brun) ;
- i, î notés i (ex. : lit, gîte).

Par ailleurs, d'autres phonèmes (les semi-voyelles) peuvent être du point de vue du contenu acoustique (celui qui nous intéresse) décomposés en plusieurs phonèmes :

wi = ou + i (ex. Louis) ;

5 ui = u + i (ex. = lui) ;

li = i + e (ex. : paille).

Le groupe 1 voit donc son effectif réduit de 24 phonèmes linguistiques à 17 phonèmes acoustiques : a, e, i, o, u, on, ou, ain, an, au, l, m, n, r, é, è, eu. Cette simplification est légitime dans la mesure où les sons regroupés ou décomposés sont phonétiquement semblables à leur substitut. Ainsi, l'analyse du contenu phonétique du mot "champignon" par exemple montre que le phonème "gn" se décompose en deux parties, une quasi-identique au "n" et l'autre au "i". L'ensemble des 36 phonèmes linguistiques se ramène donc à 29 entités phonétiques de base dont le tableau suivant reprend les caractéristiques.

	ELEMENTS	CARACTERISTIQUES
20	GROUPE 1 a, e, i, o, u, on, ou, ain, an, au, l, m, n, r, é, è, eu	énergie comprise de 0 à 5 kHz durée = 150 ... 250 msec signal quasi périodique sur toute la durée durée relativement longue
25	GROUPE 2 f, s, ch, j, z, v	énergie comprise de 0 à 12 ... 15 kHz durée = 150 ... 250 msec durée relativement longue
30	GROUPE 3 p, t, k, b, d, g	énergie comprise de 0 à 5 ... 15 kHz durée = 10 à 100 msec évolution temporelle rapide des paramètres intensité et contenu fréquentiel durée relativement courte

Il est important de préciser que les approximations ci-dessus (regroupements et décompositions) ont été introduites par souci de facilité. Elles ne constituent pas nécessairement une condition essentielle pour la mise en oeuvre de l'invention. D'autres
5 variantes d'exécution en fonction des langues en cause notamment sont possibles.

Ainsi qu'il a été indiqué, le principe sur lequel repose l'invention consiste en ce qu'une analyse fréquentielle spécifique est appliquée à chacun des groupes.

10 Ainsi pour le groupe 1, il n'est pas nécessaire d'étudier le spectre au-delà de 5 kHz puisque l'énergie s'y trouve concentrée avant. Pour le groupe 2, une analyse fréquentielle est nécessaire jusqu'à 12 kHz. Pour ces deux classes où les sons peuvent être prolongés, l'évolution temporelle des paramètres fréquents est
15 pratiquement nulle et si on calcule par exemple les spectres en amplitude d'un phonème dont la durée est prolongée pendant une seconde, on constate que l'analyse fréquentielle d'une portion de 20 ms, 50 ms, 500 ms ou de la totalité du phonème donne des résultats identiques ou similaires. L'information est redondante dans le
20 temps et une analyse fréquentielle effectuée sur une petite fenêtre prise n'importe où en dehors des zones où l'effet de voisinage est trop important, suffit pour caractériser ces sons. En ce qui concerne les plosives (groupe 3) par contre, l'information temporelle est très importante et on comprend que l'information présentée par le
25 caractère image doit donc tenir compte de l'évolution temporelle du contenu fréquentiel. L'information n'est pratiquement pas redondante dans le cas des plosives. Il est donc avantageux d'exploiter tout ou une partie notable du signal pour l'identification de la plosive. L'information présentée sera la même ou de même
30 nature ou équivalente à l'information présentée par un sonagramme en particulier dans le sens où elle exploite une partie notable du signal temporel.

La nature "hétéroclite" de cet alphabet phonétique naturel est justifiée par l'existence de trois groupes de sons qui lui
35 donnent naissance.

Le principe étant admis, on peut utiliser avantageusement un "alphabet" dans lequel les sons du groupe 1

sont représentés par la courbe enveloppe du spectre d'énergie
jusque 5 kHz en couleur par exemple rouge, ceux du groupe 2 en
couleur verte et les derniers par une présentation avantageusement
en couleur différente des groupes 1 et 2, du type sonagramme ou
d'un autre type de représentation par caractère-image contenant une
information de même type que dans la représentation sonagramme.
On peut bien entendu procéder à une discrimination selon d'autres
critères que la couleur, en particulier des critères géométriques
des caractères-images.

L'analyse fréquentielle de la quasi-totalité du signal
temporel de la plosive ou du moins d'une partie notable de ce signal
ou même de la totalité du signal temporel de la plosive plus une
partie du signal temporel du phonème précédent et/ou suivant peut
être avantageusement considérée comme un tel caractère-image.

Un traitement préalable nécessaire doit remplir deux
fonctions : d'une part, découper la parole en phonèmes et d'autre
part, décider à quel groupe (1, 2 ou 3) appartiennent ces phonèmes.

La première fonction, la segmentation est présente dans
tout système de reconnaissance analytique ; elle ne présente en
général aucune difficulté. A titre d'exemple, un segmenteur
utilisant des informations simples comme le nombre de passages à
zéro (NPZ) de l'amplitude, détecte 99 % des phonèmes (ARPA).

La seconde étape ne pose pas de problèmes puisque les
caractéristiques fréquentielles et/ou temporelles sont différentes
d'une classe à l'autre.

EXEMPLES D'APPLICATION DE L'INVENTION

La figure 2 représente l'évolution temporelle de la
pression en fonction du temps (3 secondes) pour la phrase "Alain est
mon ami".

Cette phrase est constituée de 10 phonèmes que l'on
aperçoit très distinctement dans le signal temporel de la figure 2,
ceux-ci étant mis en évidence par les fenêtres en noir.

Dans ce premier exemple ayant choisi comme caractère-
image pour représenter le phonème, la courbe de la prédiction
linéaire (enveloppe de la transformée de Fourier) représentée avec
comme abscisse une échelle linéaire des fréquences allant de 0 à 5
kHz et une échelle logarithmique des ordonnées (décibels), on a
représenté l'écriture de la phrase-exemple sur la figure 3.

On voit que chaque phonème est représenté par un caractère-image et un seul. Le caractère-image est présenté immobile précédé et/ou suivi par un certain nombre d'autres caractères-images. On découvre d'une manière qui paraît instantanée la totalité des caractères-images présentés.

Un lecteur ayant subi un apprentissage peut lire chacun des phonèmes représenté par les caractères-image de la figure 3. Les fenêtres identifiées par les chiffres de 1 à 10 sur la figure 2 correspondent à des zooms de 200 ms qui sont repris successivement dans les figures 4, 6 à 14.

Elles permettent d'illustrer le caractère redondant de l'information dans le signal temporel et illustre la façon avec laquelle les caractères-images représentant les phonèmes ont été choisis.

La figure 4 présente en haut le premier zoom du signal temporel qui correspond au premier phonème "A" de la phrase et décrit en-dessous l'évolution de son spectre caractéristique en amplitude calculé sur des fenêtres de 20 ms désignées par les lettres minuscules a jusque f. Une première constatation s'impose ; l'allure du spectre est suffisamment stable sur l'ensemble des 6 fenêtres de calcul et reste conforme au modèle du phonème isolé A tel qu'on peut le voir sur la figure 5 qui représente les caractères-images correspondants à des phonèmes du groupe 1 émis de façon isolée par le même locuteur que celui de la phrase concernée.

Le signal temporel de la figure 6 contient l'entièreté du son "L" - mis en évidence par les spectres relatifs aux fenêtres a-b-c et la transition "L" - "AIN" caractérisée par les fenêtres d-e-f.

On remarque en effet dans ces dernières, l'apparition de l'allure générale du "AIN" qui reste toutefois influencée par le relèvement significatif du "L" à 1,5 kHz. Il s'agit d'un effet de voisinage.

La figure 7 quant à elle montre par ses fenêtres a-b-c-d-e un modèle du "AIN" plus typique.

La fenêtre a de la figure 8 cloture le mot ALAIN et la fenêtre b représente le début du phonème "è".

À la figure 9, on reconnaît aux fenêtres a-b-c-d le son "è" et on devine aux fenêtres e-f l'apparition timide du "M".

Les courbes a et b de la figure 10 sont à ce propos beaucoup plus révélatrices quant à la présence du "M" de même que le sont les courbes e et f pour le "ON".

La figure 10 montre aux fenêtres b-c-d-e-f des exemples de spectres du son "N".

La figure 11 est relative au phonème "A".

La figure 12 est relative au phonème "M".

5 La figure 13 est relative au phonème "I".

Par éducation, il est possible de s'éduquer à lire les caractères images tels ceux représentés à la figure 3. Des essais ont montré que les phonèmes isolés du groupe 1 prononcés par un locuteur auquel on s'est habitué (c'est-à-dire que l'on s'est entraîné à reconnaître les caractères-images correspondant aux phonèmes isolés du groupe 1 émis par ce locuteur) sont reconnus dans une proportion de 99 %. Quant aux phonèmes coarticulés dans le langage, ils sont identifiés dans une proportion de 96 % ; le sens du mot et de la phrase permet de lever ultérieurement l'ambiguïté résiduelle comme dans la lecture d'une écriture manuscrite. On parle de phonème coarticulé lorsque le phonème n'est pas produit isolé mais dans un mot du langage parlé.

15 Les figures 15, 16, 17, 18, 19, 20, 21, 22 reprennent la même démarche pour le mot "sosie", phonétiquement sozie.

20 La figure 17 représente pour un même locuteur les caractères-images correspondant au groupe 2 de phonèmes, les fricatives. Pour les caractères-images correspondant aux fricatives, l'échelle des fréquences a été choisie s'étendant de 0 kHz à 12,5 kHz.

REVENDICATIONS

1. Procédé de reconnaissance de la parole et en particulier d'écriture de la parole comportant une analyse fréquentielle des phonèmes émis par un locuteur, caractérisé en ce que les phonèmes subissent un traitement d'analyse fréquentielle
5 spécifique qui génère une image statique pour chaque phonème présentée immobile pendant un temps suffisant pour permettre la reconnaissance du phonème concerné.

2. Procédé selon la revendication 1 caractérisé en ce que
10 l'image statique représentant le phonème apparaisse d'une façon paraissant instantanée sur l'écran.

3. Procédé selon les revendications 1 ou 2 caractérisé en ce qu'il comporte un affichage simultané de plusieurs phonèmes successifs.

4. Procédé selon l'une quelconque des revendications 1 à 3 caractérisé en ce que les images forment des caractères ou sont représentées par des caractères qui sont groupés en suite successive dans un mot que l'on sépare utilement par un espacement des caractères relatifs à des mots différents.
15

5. Procédé selon l'une quelconque des revendications 1 à 4 caractérisé en ce qu'il comporte l'affichage de la longueur de l'émission de l'élément verbal, de l'évolution dans le temps de son amplitude et de la régularité de l'émission acoustique dans l'élément verbal, dans le mot et dans la phrase, ainsi
20 que les caractéristiques de périodicité de chaque phonème.

6. Procédé selon l'une quelconque des revendications 1 à 5 caractérisé en ce qu'il comporte une opération d'analyse fréquentielle spécifique des phonèmes selon qu'ils appartiennent à un premier groupe constitué par les phonèmes autres que les
30 consonnes fricatives ou les consonnes plosives, à un deuxième groupe constitué par les consonnes fricatives et à un troisième groupe constitué par les consonnes plosives.

7. Procédé selon l'une quelconque des revendications 1 à 6 caractérisé en ce que l'on procède à une analyse fréquentielle du premier groupe au moins jusqu'à une valeur de l'ordre de 4 à 5 kHz sur une durée de l'ordre d'au moins une quasi-période.
35

8. Procédé selon la revendication 7 caractérisé en ce que l'on procède à une analyse fréquentielle du deuxième groupe au

moins jusqu'à une valeur de l'ordre de 12 kHz pendant une durée supérieure à 1 milliseconde, et ne dépassant pas la durée du phonème.

5 9. Procédé selon la revendication 7 ou 8 caractérisé en ce que l'on procède à une analyse fréquentielle du troisième groupe au moins jusqu'à 5 kHz pendant une durée suffisante pour tenir compte de l'évolution temporelle des caractéristiques fréquentielles.

10 10. Procédé selon l'une quelconque des revendications 1 à 9 caractérisé en ce que des signaux analogiques captés par un microphone subissent une digitalisation et sont soumis à l'aide d'un algorithme à une analyse fréquentielle propre au groupe afin d'obtenir une écriture de la parole basée sur l'analyse fréquentielle de celle-ci, analyse qui fournit un alphabet phonétique naturel.

15 11. Procédé selon l'une quelconque des revendications 6 à 10 caractérisé en ce que l'écriture correspondant à chacun des 3 groupes précités subit une représentation particulière pour chaque groupe, notamment une représentation de couleur différente pour le groupe 1 de la couleur retenue pour la représentation du groupe 2
20 tandis que les phonèmes du groupe 3 subissent une représentation du type sonagramme.

25 12. Procédé selon l'une quelconque des revendications 6 à 10 caractérisé en ce que l'écriture correspondant à chacun des 3 groupes précités subit une représentation de couleur différente pour le groupe 1 de la couleur retenue pour le groupe 2 tandis que les phonèmes du groupe 3 subissent une représentation de couleur
différente des phonèmes des groupes 1 et 2.

30 13. Procédé selon l'une quelconque des revendications 6 à 12 caractérisé en ce que l'écriture correspondant à chacun des 3 groupes précités présente comme caractère-image correspondant à chaque phonème, pour le groupe 1 la transformée fréquentielle obtenue en exploitant une partie seulement ou la totalité du signal temporel du phonème et ce au moins jusqu'à environ 4 à 5 kHz, pour
35 le groupe 2, la transformée fréquentielle obtenue en exploitant une partie seulement ou la totalité du signal temporel du phonème et ce au moins jusqu'à environ 12 kHz, pour le groupe 3, la transformée fréquentielle obtenue en exploitant soit la totalité de l'information acoustique du signal temporel du phonème, soit la quasi-totalité de

l'information acoustique du signal temporel du phonème, soit la
totalité de l'information acoustique du signal temporel du phonème
avec une partie de celle d'un ou de deux phonèmes voisins, soit une
partie de l'information acoustique du signal temporel du phonème
5 avec une partie de celle d'un phonème voisin.

14. Procédé selon l'une quelconque des revendications 6 à
13 caractérisé en ce que l'écriture correspondant à chacun des
groupes précités subit une représentation particulière pour chaque
groupe selon des formes géométriques différentes pour chaque
10 groupe.

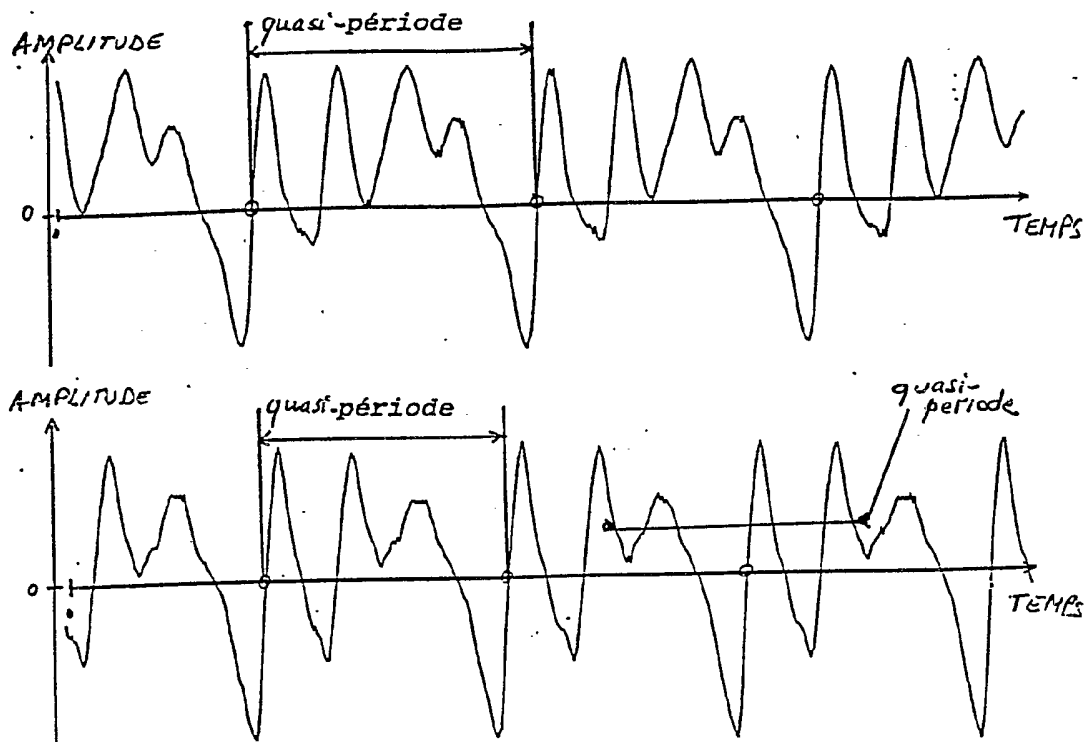


FIG 1

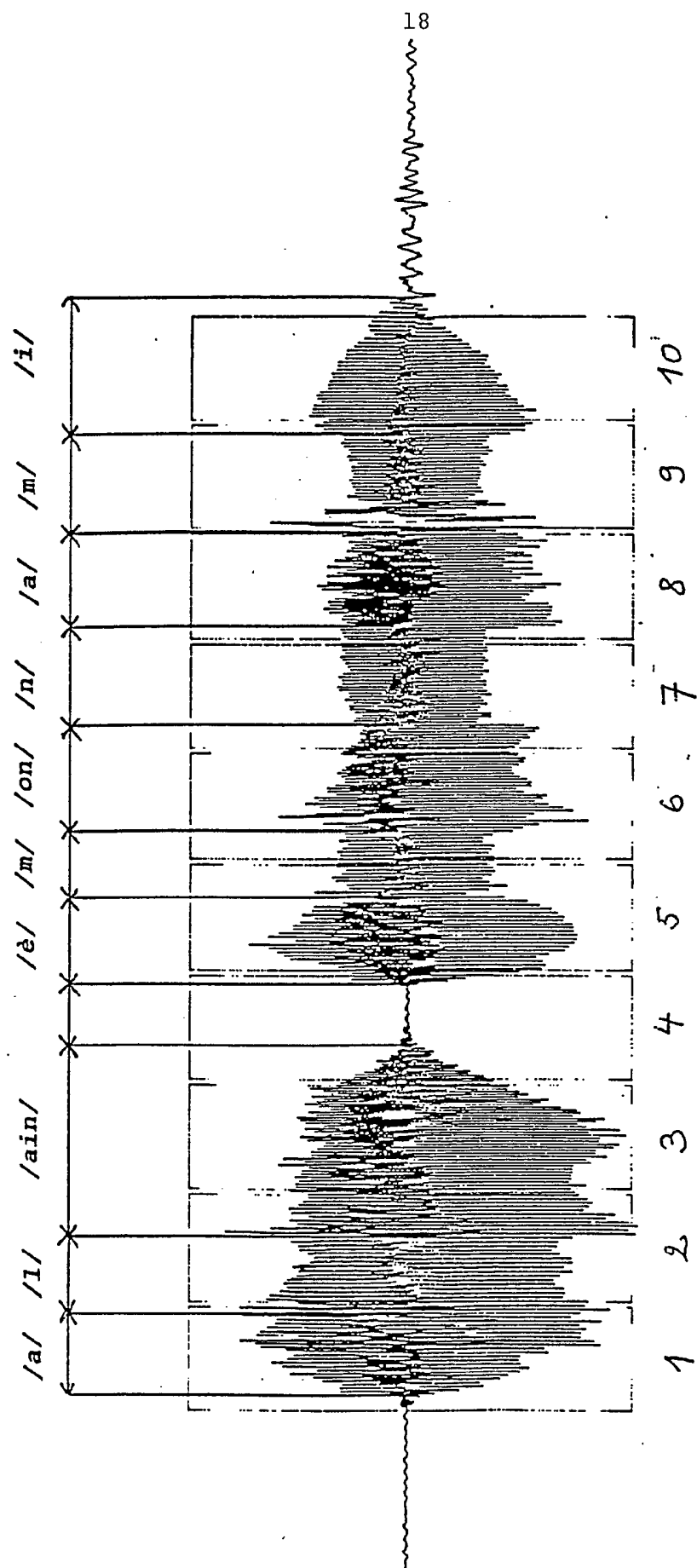


FIG 2

19

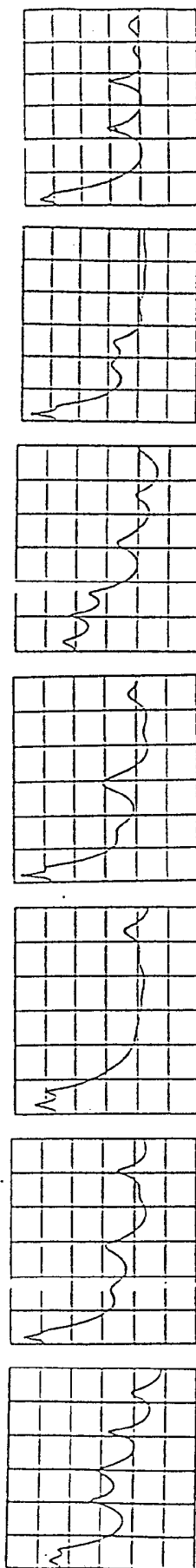
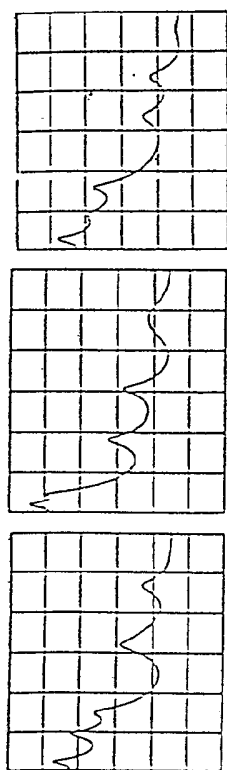
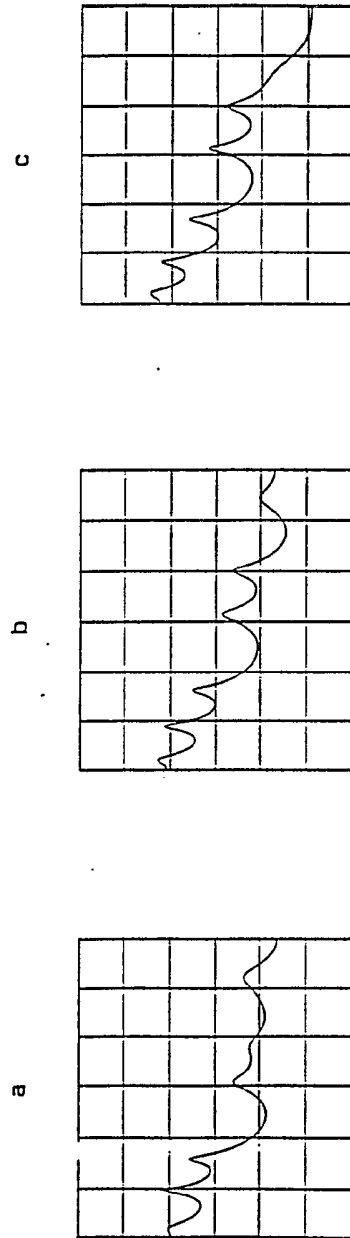
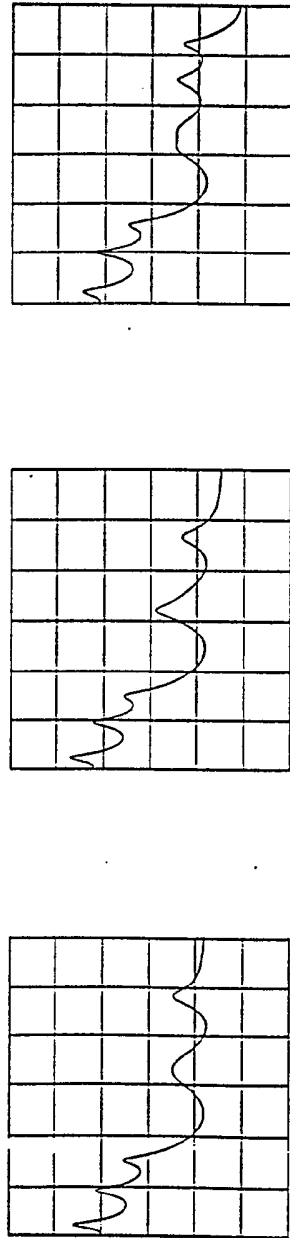
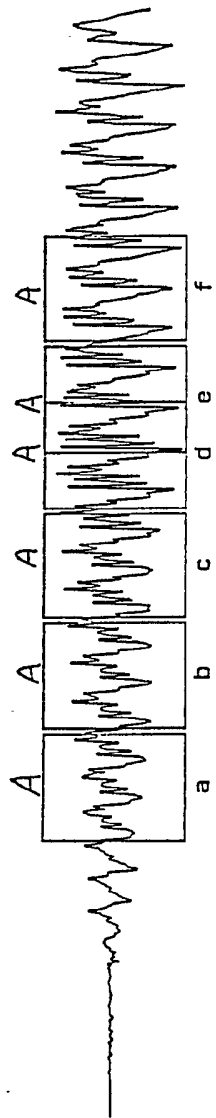


FIG 3



f

e
fenetre 1
FIG 4

d

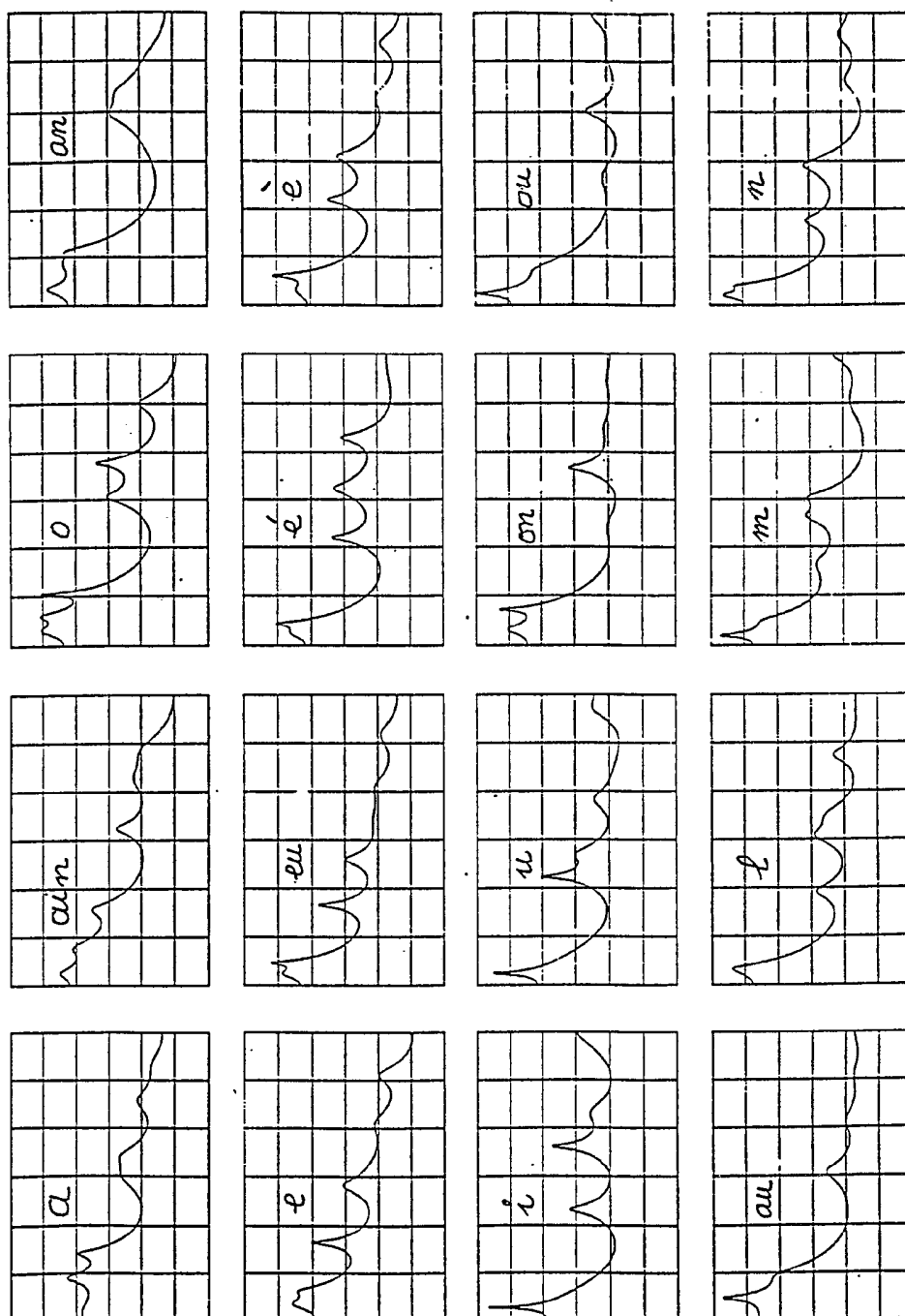
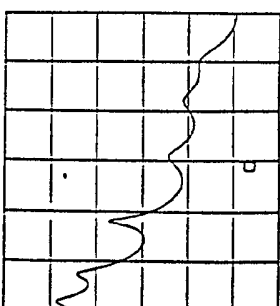
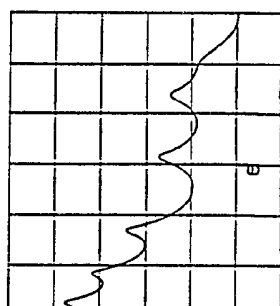
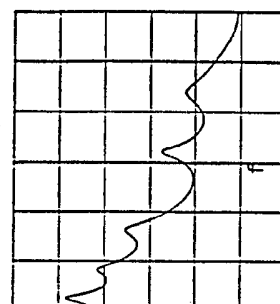
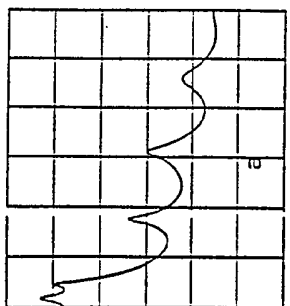
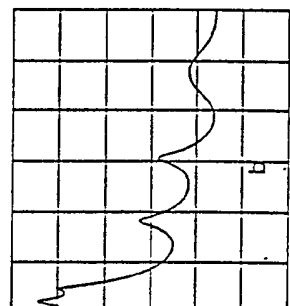
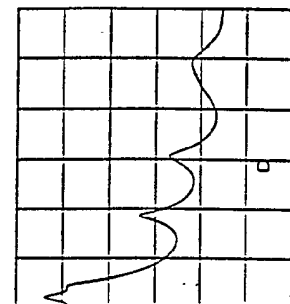
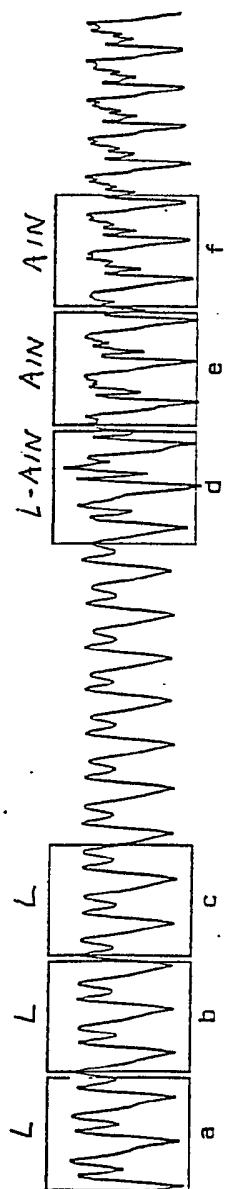
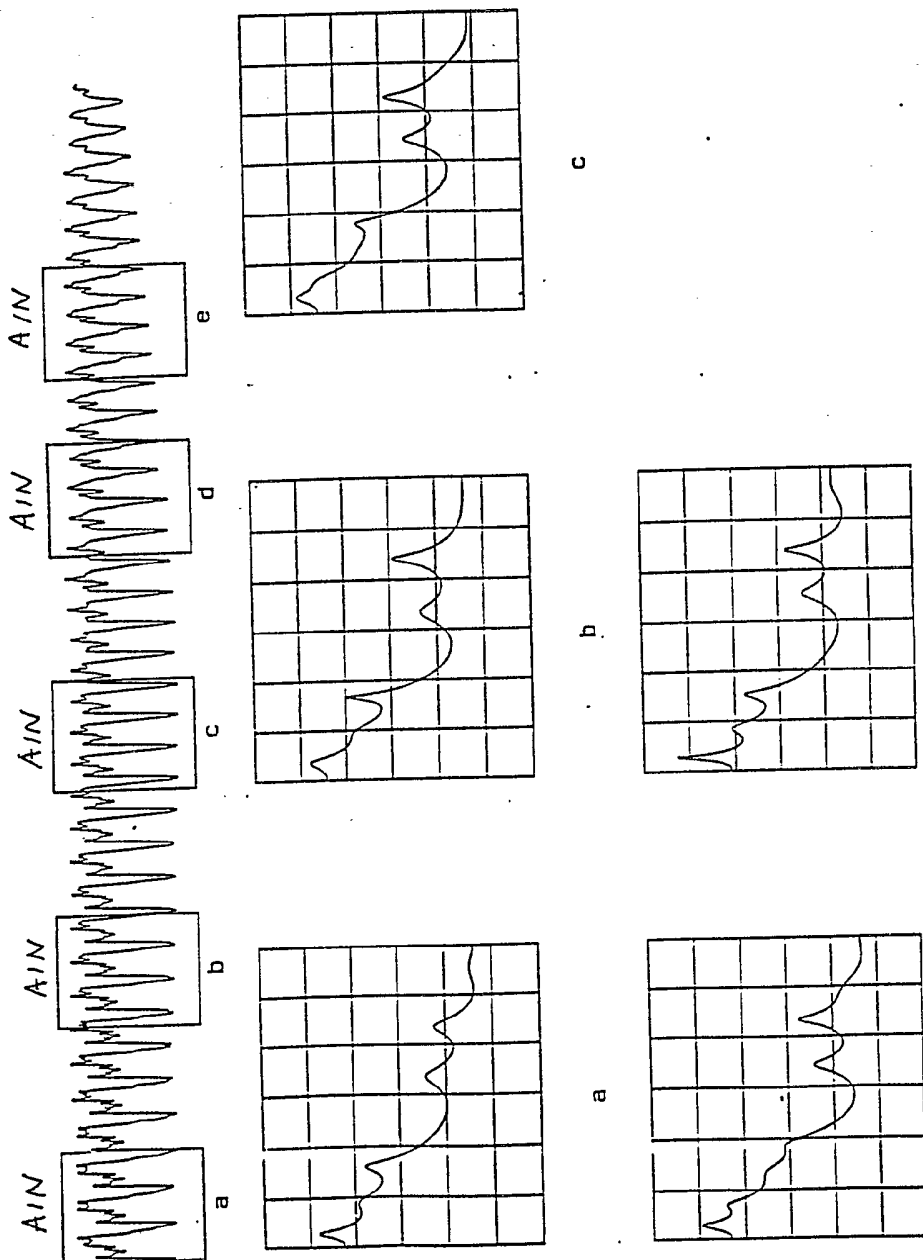


FIG 5

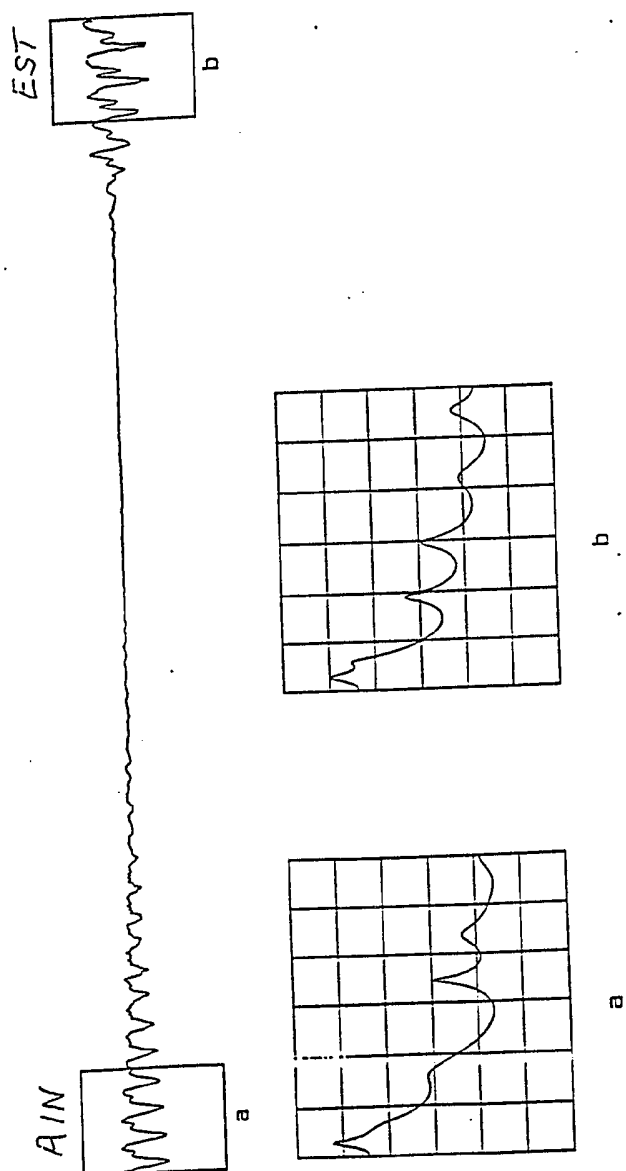


fenetre 2

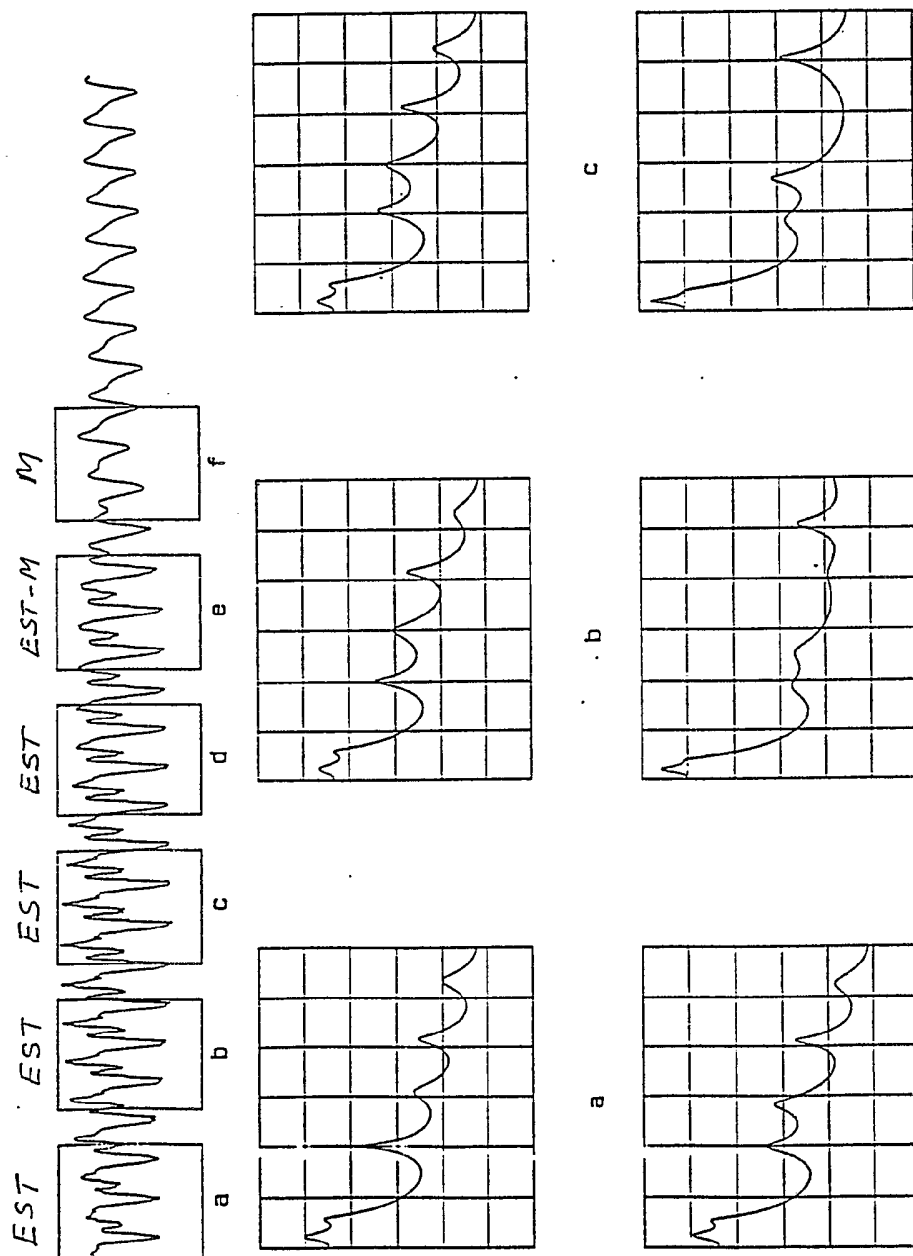
FIG 5



^e
fenetre 3
FIG 7



fenetre 4
FIG 8



f

fenetre 5

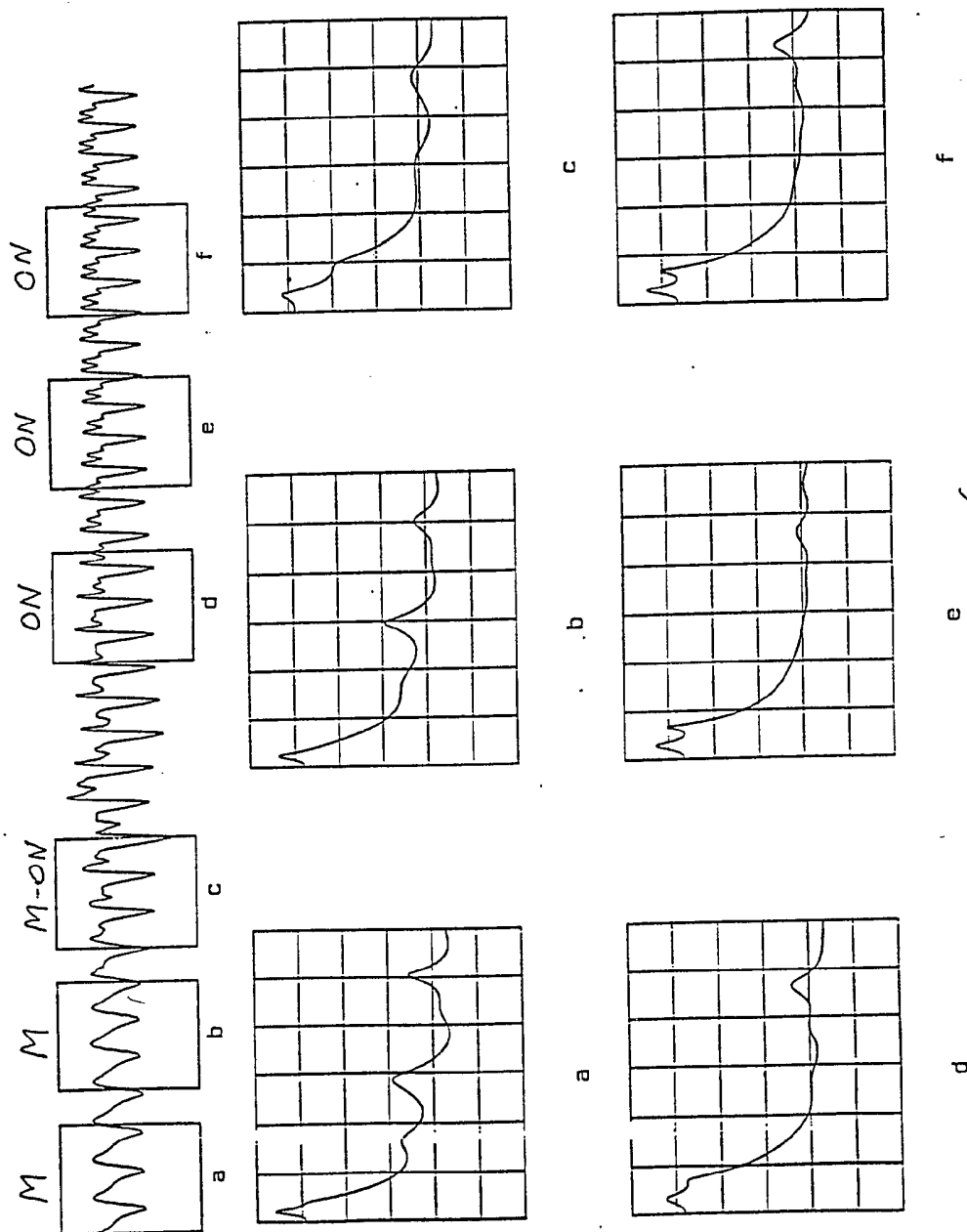
FIG 9

d

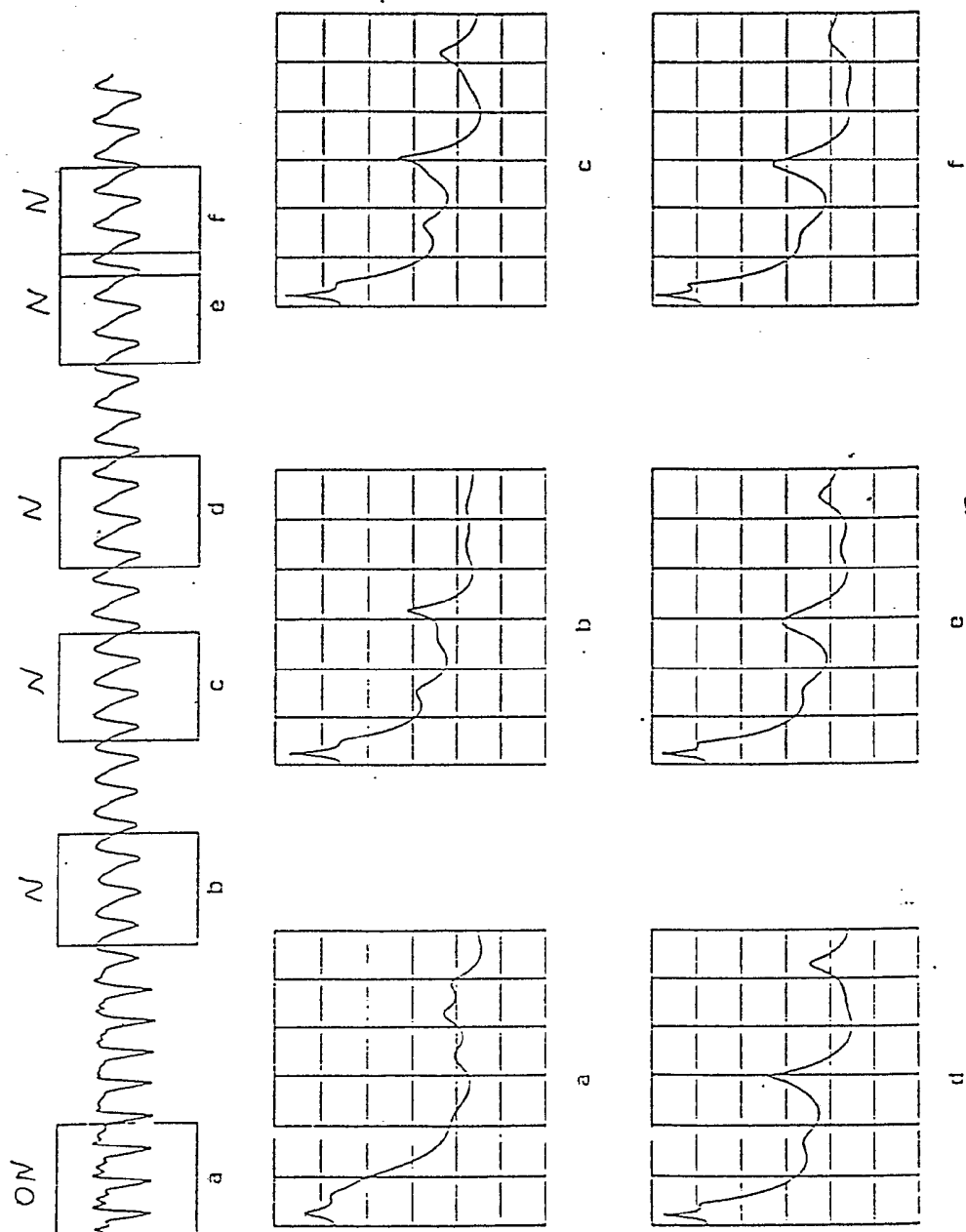
c

b

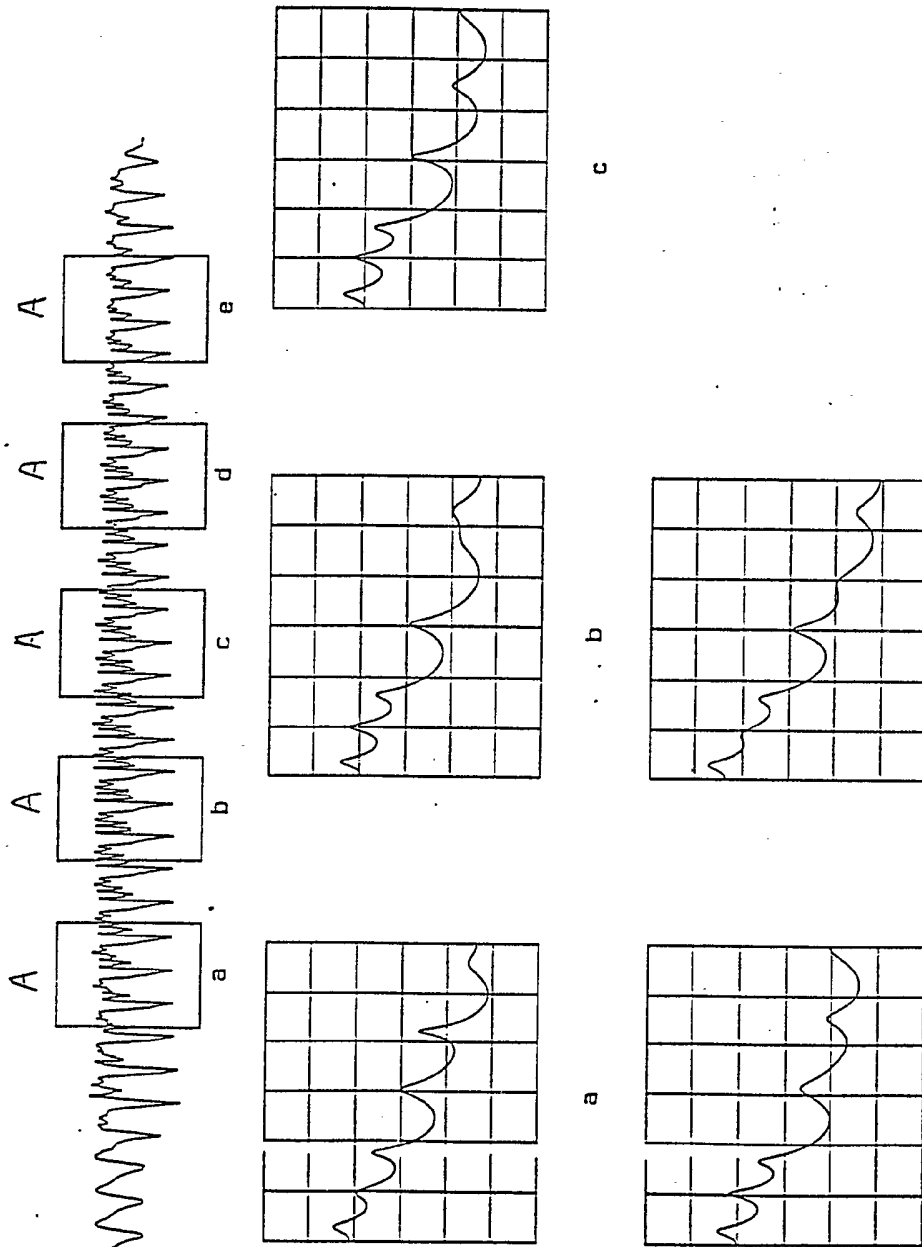
a



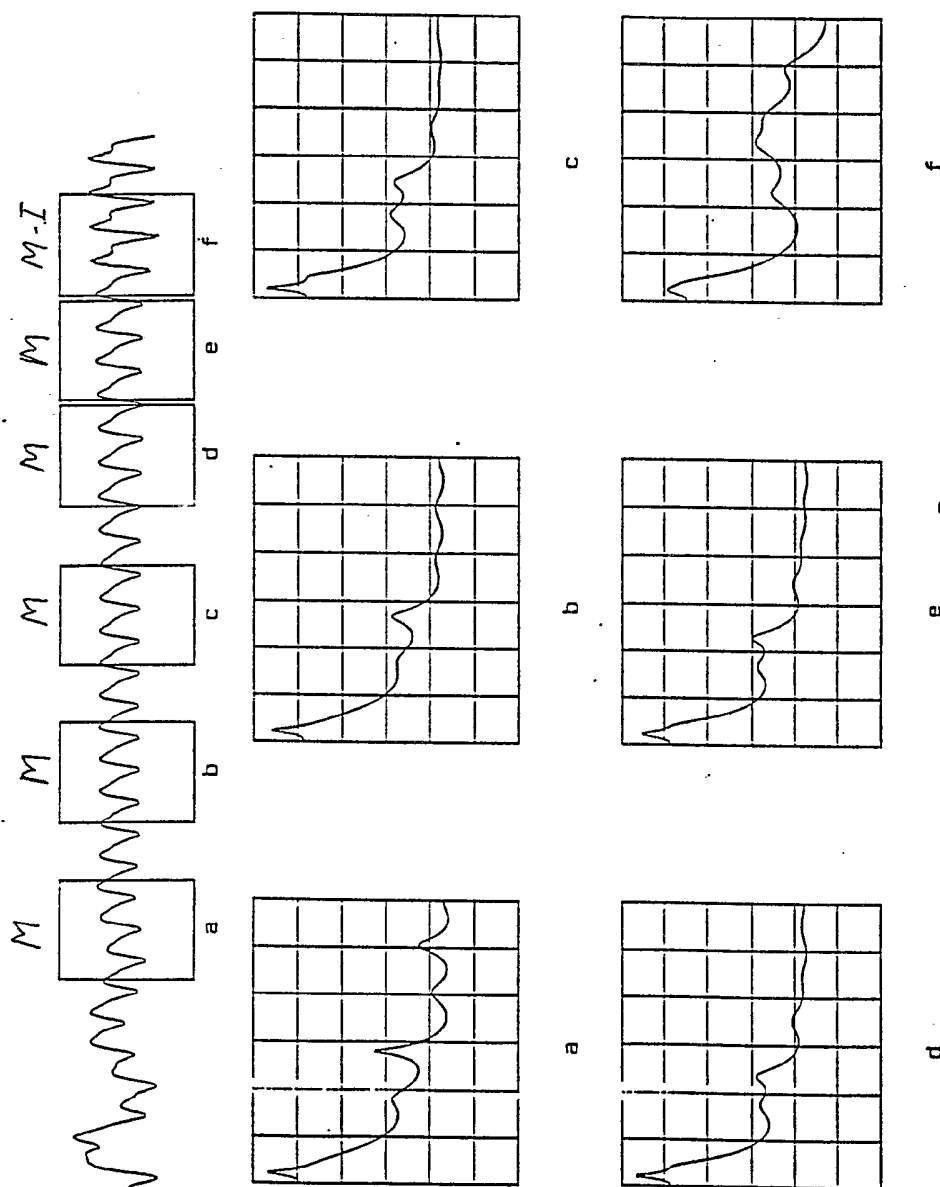
fenetre 6
FIG 10



fenetre 7
FIG 11



fenetre 8
FIG 12



e
fenetre 9
FIG 13

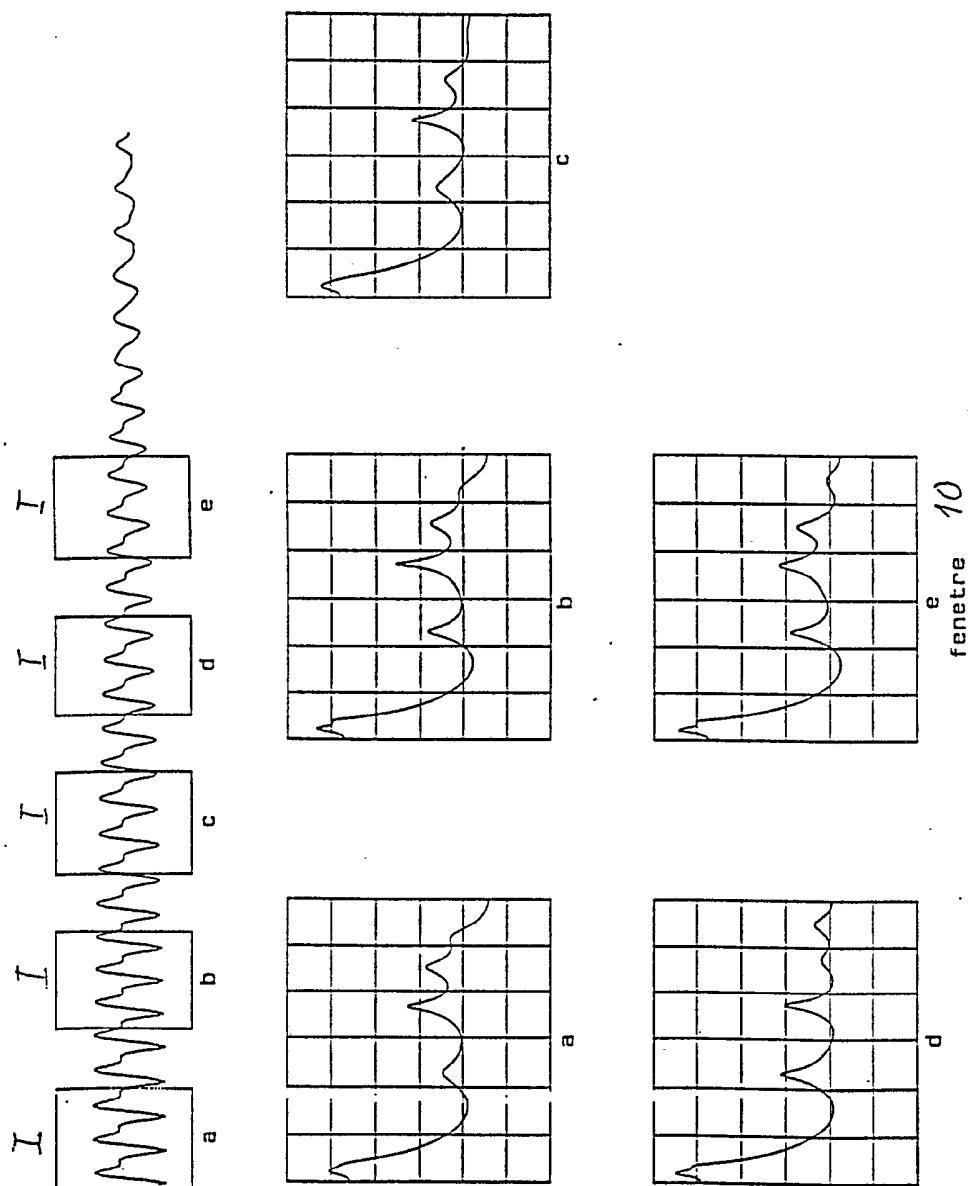


FIG 14

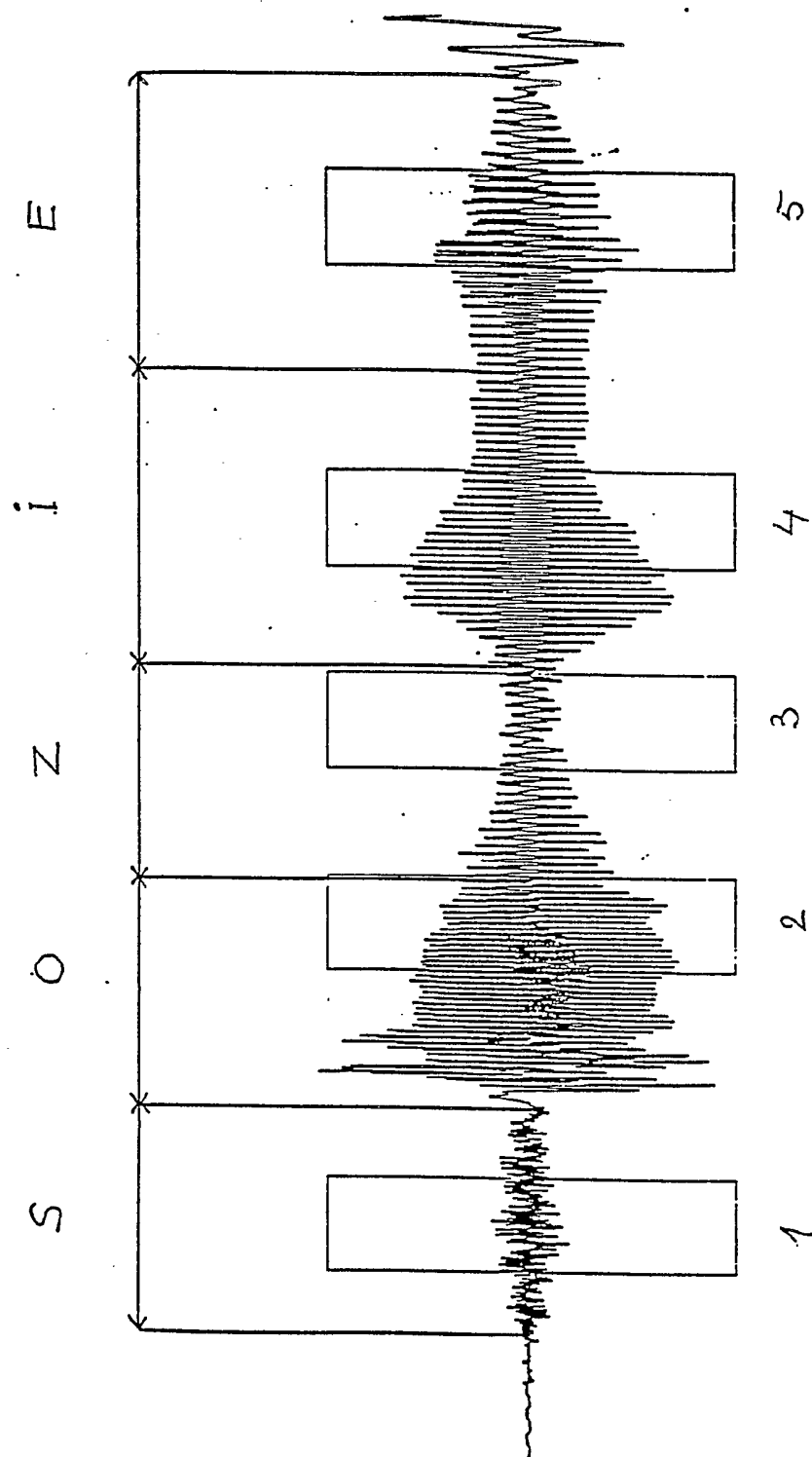


FIG 15

32

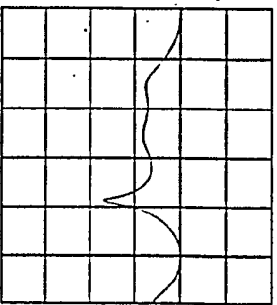
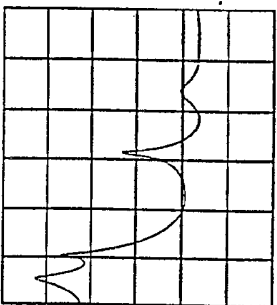
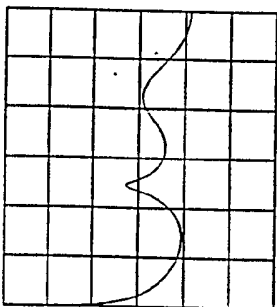
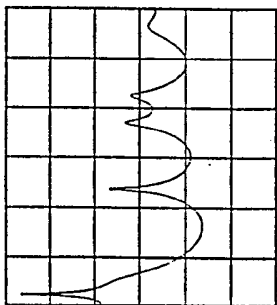
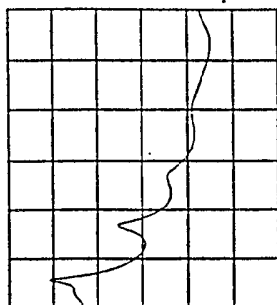
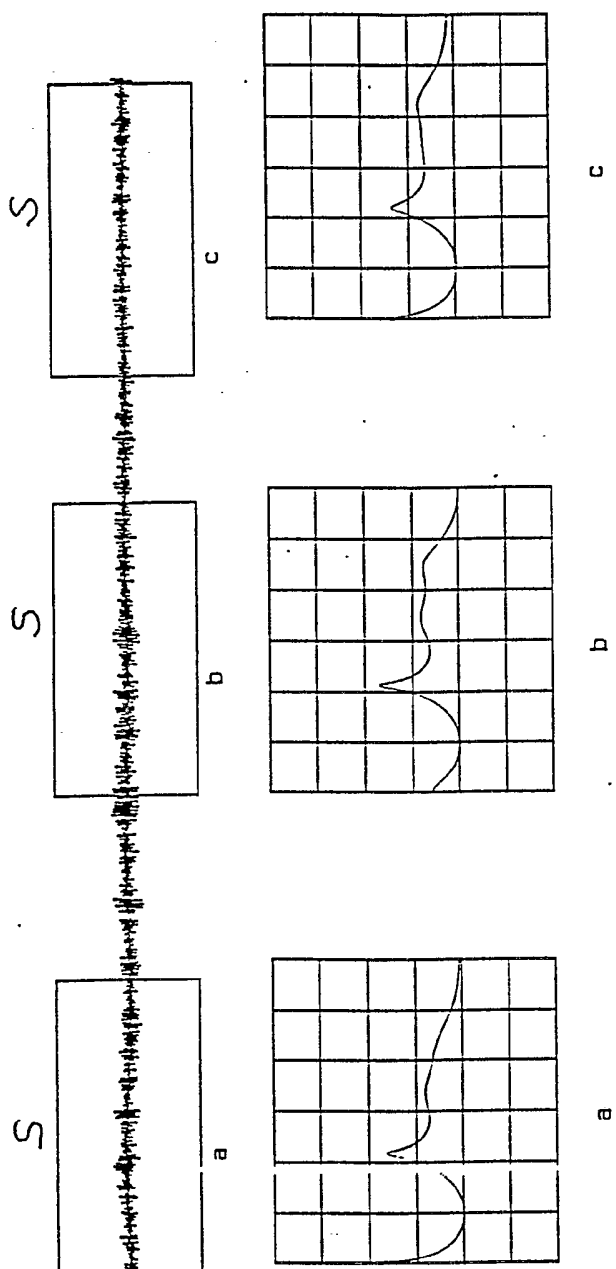


FIG 16



fenetre 1
FIG 17

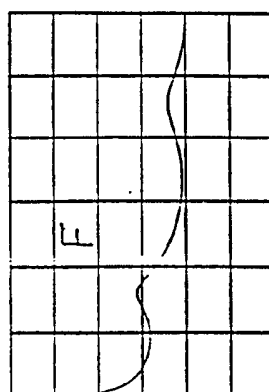
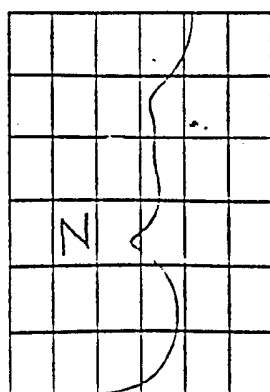
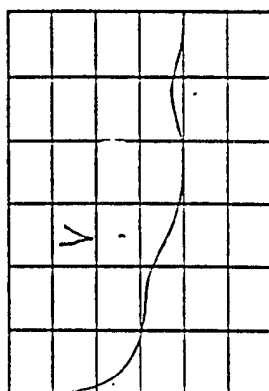
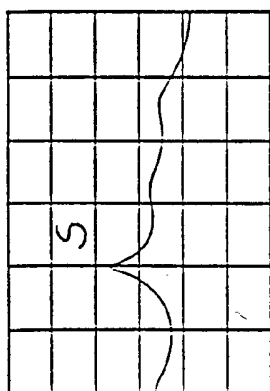
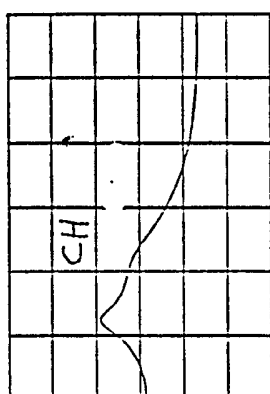
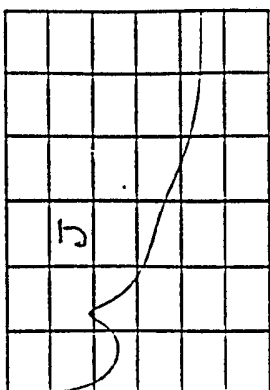
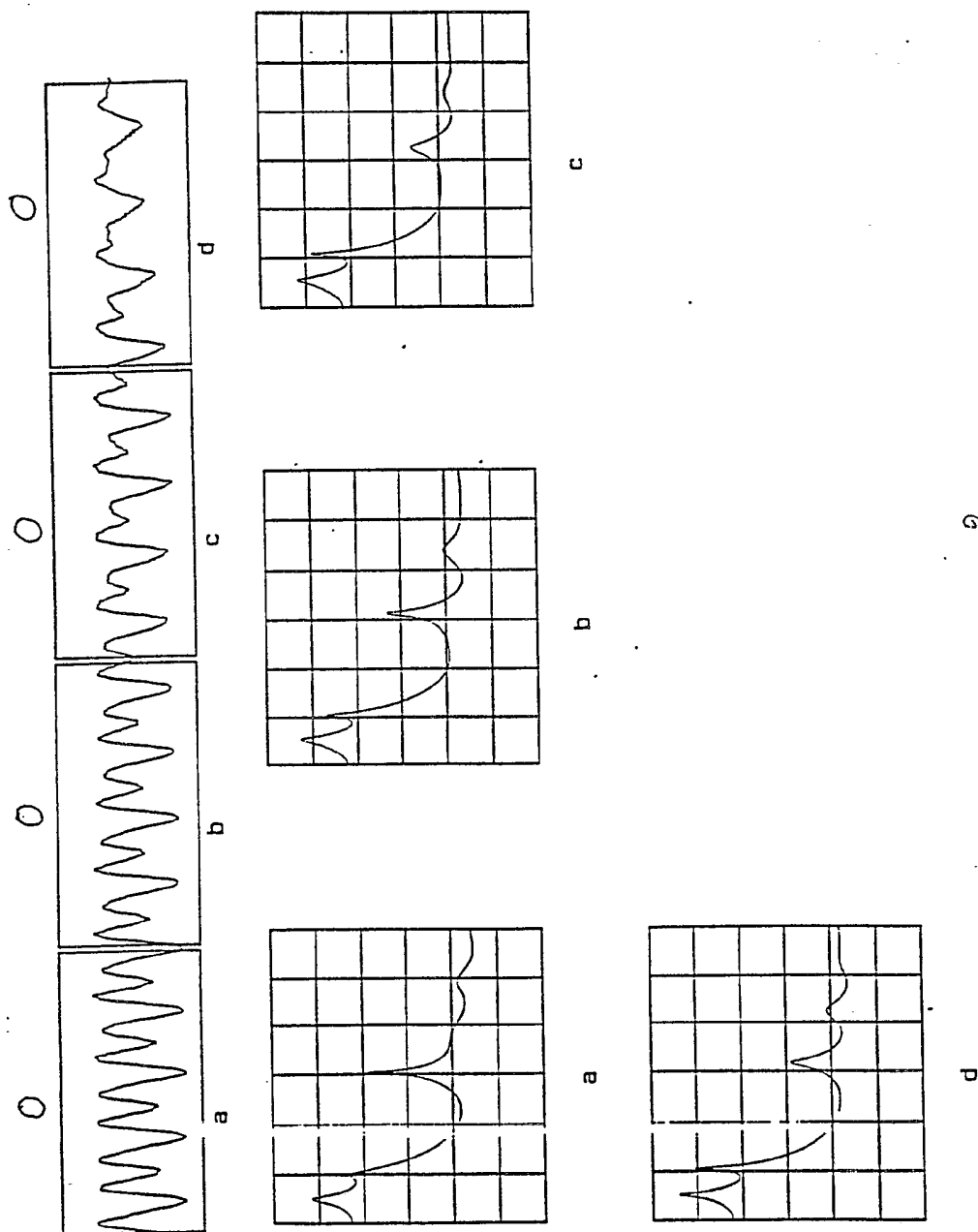
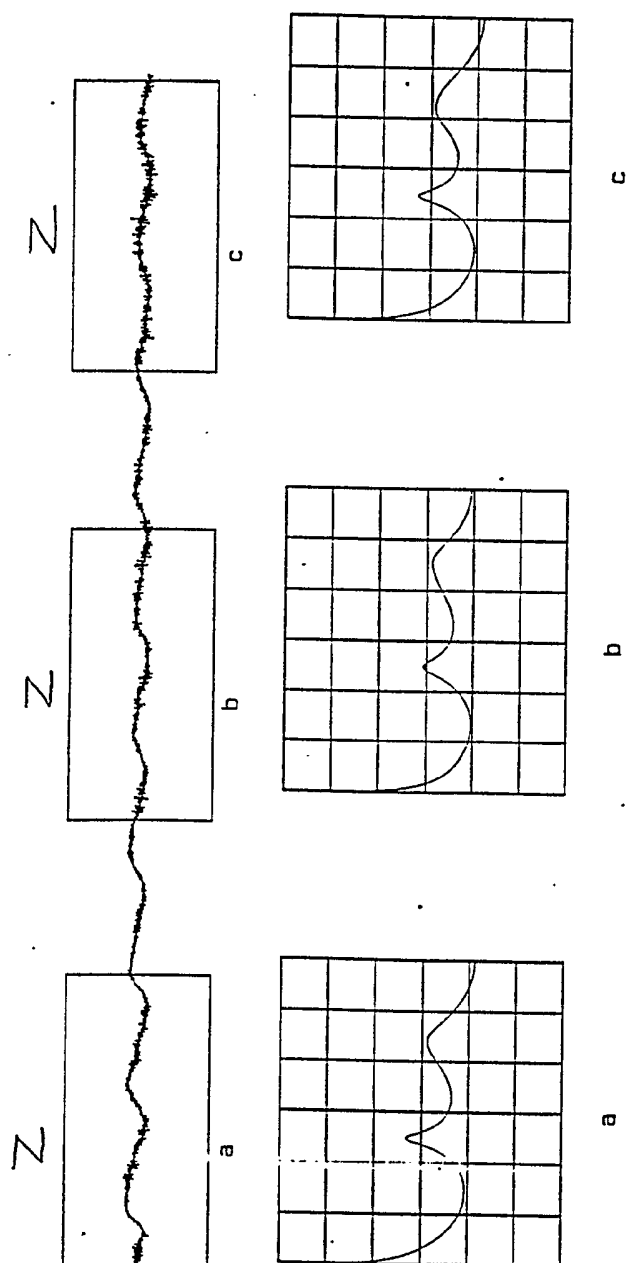


FIG 18

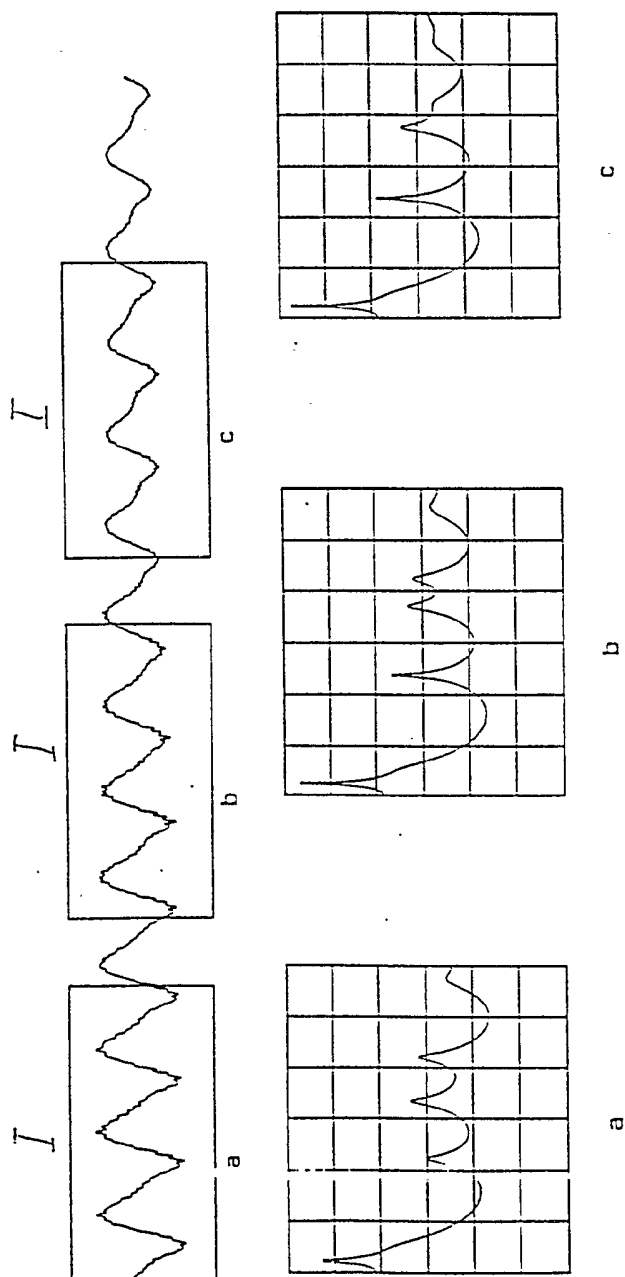


fenetre 2
FIG 19

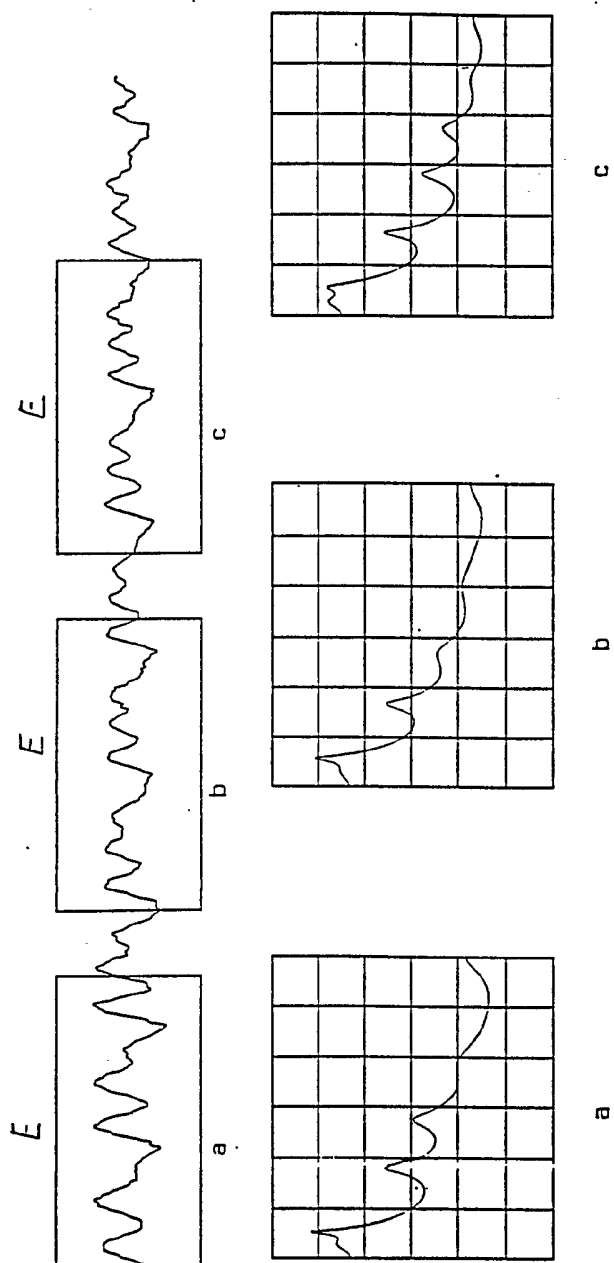


fenetre 3

FIG 20



fenetre 4
FIG 21



fenetre 5

FIG 22



Office européen
des brevets

RAPPORT DE RECHERCHE

établi en vertu de l'article 21 § 1 et 2
de la loi belge sur les brevets d'invention
du 28 mars 1984

Numero de la demande
nationale

BE 8800993

BO 1409

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int. Cl.4)
X	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. ASSP-33, no. 1, février 1985, pages 164-173, IEEE, New York, US; A. WATANABE et al.: "Color display system for connected speech to be used for the hearing impaired" * Résumé; page 165, colonne de gauche, lignes 13-18 * ---	1,2,11,12	G 10 L 7/10
X	PROCEEDINGS OF THE NATIONAL ELECTRONICS CONFERENCE, vol. 32, octobre 1978, pages 205-207, Oak Brook, US; R.M. CAMPBELL: "Systems for aiding the speech and hearing of the deaf" * Page 205, colonne de droite, lignes 24-39; page 206, colonne de gauche, lignes 1-5 * ---	1,2	
X	US-A-4 284 846 (J. MARLEY) * Colonne 4, ligne 16 - colonne 5, ligne 35 * ---	1-4	DOMAINES TECHNIQUES RECHERCHES (Int. Cl.4)
A	IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS, vol. AU-16, no. 2, juin 1968, pages 227-234, New York, US; J.M. PICKETT: "Recent research on speech analyzing aids for the deaf" * Page 230, colonne de gauche, lignes 15-20; colonne de droite, lignes 27-32 * --- -/-	1,5	G 01 L 7/10
Date d'achèvement de la recherche		Examineur	
02-05-1989		ARMSPACH J. F. A. M.	
CATEGORIE DES DOCUMENTS CITES		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons ----- & : membre de la même famille, document correspondant	
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire			

ABSENCE D'UNITE D'INVENTION

La présente demande ne satisfait pas à l'exigence relative à l'unité d'invention et concerne plusieurs inventions ou pluralités d'inventions, à savoir

1. Revendications :

2. Revendications :

3. Revendications :

4. Revendications :

Le présent rapport de recherche a été établi de façon complète pour les parties de la demande qui se rapportent à l'invention ou pluralité d'inventions mentionnée dans les revendications:

ETENDUE DE LA RECHERCHE

Compte tenu des documents considérés comme pertinents, le présent rapport de recherche a été établi de façon complète pour les parties de la demande qui se rapportent à l'invention ou pluralité d'inventions mentionnée en premier lieu dans les revendications, à savoir les revendications :

Les éléments figurant dans les

1. Revendications :

2. Revendications :

3. Revendications :

4. Revendications :

n'ont pas été pris en considération que dans le cadre de la recherche relative aux caractéristiques de l'invention ou de la pluralité d'inventions mentionnée en premier lieu dans les revendications



Office européen
des brevets

RAPPORT DE RECHERCHE

établi en vertu de l'article 21 § 1 et 2
de la loi belge sur les brevets d'invention
du 28 mars 1984

Numero de la demande
nationale
Page 2

BE 8800993
BO 1409

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int. Cl.4)
A	IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS, vol. AU-20, no. 1, mars 1972, pages 16-23, New York, US; A.J. GOLDBERG: "A visual feature indicator for the severely hard of hearing" * Page 19: "Display operation" * ---	1,2,5,6	DOMAINES TECHNIQUES RECHERCHES (Int. Cl.4)
A	IEEE TRANSACTIONS ON COMPUTERS, vol. C-20, no. 9, septembre 1971, pages 1057-1063, New York, US; J. MAKHOUL: "Speaker adaptation in a limited speech recognition system" * Figure 2 * ---	1,2,4	
A	DE-B-1 214 016 (TELEFUNKEN PATENTVERWERTUNGS GmbH) * Colonne 1, lignes 27-52 * ---	1,2	
A	EP-A-0 108 609 (ING. C. OLIVETTI & C.) * Résumé * ---	1,2,4,5	
A	THE JOURNAL OF THE ACOUSTICAL SOCIETY OF JAPAN, vol. 31, no. 3, mars 1975, pages 189-195, Tokyo, JP; Y. ISHIDA et al.: "On a speech trainer for deaf persons" * Résumé * ---	1,2,5,6	
A	US-A-3 463 885 (H.W. UPTON) * Colonne 3, ligne 40 - colonne 4, ligne 61 * ---	1,2,6-9,11,13,14	
A	EP-A-0 050 443 (RESEARCH TRIANGLE INSTITUT) * Résumé; figure 4 * -----	1,2,6-9,11,13,14	
Date d'achèvement de la recherche		Examineur	
02-05-1989		ARMSPACH J.F.A.M.	
CATEGORIE DES DOCUMENTS CITES			
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

**ANNEXE AU RAPPORT DE RECHERCHE
RELATIF A LA DEMANDE DE BREVET BELGE NO.**

BE 8800993
B0 1409

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche visé ci-dessus.
Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 18/05/89
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US-A- 4284846	18-08-81	GB-A, B 2020467	14-11-79
		FR-A, B 2425689	07-12-79
		DE-A, C 2918533	15-11-79
		US-A- 4181813	01-01-80
		JP-A- 55000585	05-01-80

DE-B- 1214016		Aucun	

EP-A- 0108609	16-05-84	JP-A- 59101700	12-06-84
		US-A- 4703504	27-10-87

US-A- 3463885	26-08-69	Aucun	

EP-A- 0050443	28-04-82	JP-A- 57096400	15-06-82
		AU-A- 7621381	22-04-82
		CA-A- 1182690	19-02-85
		AU-B- 551649	08-05-86
