

⑫

DEMANDE DE BREVET EUROPEEN

⑳ Numéro de dépôt: **89420197.9**

⑤① Int. Cl.⁴: **G 10 L 9/10**

㉑ Date de dépôt: **08.06.89**

③⑩ Priorité: **14.06.88 FR 8808255**

④③ Date de publication de la demande:
20.12.89 Bulletin 89/51

⑥④ Etats contractants désignés: **DE FR GB IT NL**

⑦① Demandeur: **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE**
15, quai Anatole France
F-75007 Paris (FR)

⑦② Inventeur: **Mrayati, Mohamed**
83, Chatti Amara
Damas (SY)

Carre, René
11bis, Rue Elysée Reclus
F-38100 Grenoble (FR)

Guerin, Bernard
Le Deiard
F-38430 Saint Jean De Moirans (FR)

⑦④ Mandataire: **de Beaumont, Michel**
1bis, rue Champollion
F-38000 Grenoble (FR)

⑤④ Procédé et dispositif d'analyse, synthèse, codage de la parole.

⑤⑦ Il est décrit un moyen d'analyse et de synthèse de la parole utilisant la simulation du comportement acoustique d'un tube divisé en tronçons de section variable. Les variations de section des différents tronçons d'un tube permettent d'engendrer des phonèmes lorsqu'une source de débit et pression d'air est placée analogue aux cordes vocales humaines. Par simulation on peut engendrer ces phonèmes sous forme de signaux électriques fournis à un haut-parleur. L'invention porte sur le choix des longueurs des tronçons de tube et lie ce choix à la finesse de l'approximation qu'on veut faire. Pour une approximation à trois formants (les formants sont les fréquences de résonance du tube), on divise le tube en huit tronçons de longueurs successives $L/10$, $L/15$, $2L/15$, $3L/15$, $3L/15/2L/15$, $L/15$ et $L/10$; L est la longueur totale du tube.

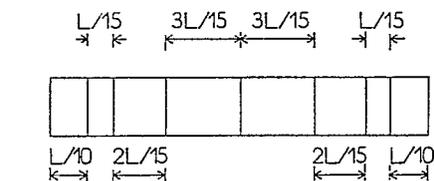


Fig.6

Description

PROCEDE ET DISPOSITIF D'ANALYSE, SYNTHÈSE, CODAGE DE LA PAROLE

L'invention concerne l'analyse, la synthèse et le codage de la parole.

Les procédés d'analyse, de synthèse et de codage de la parole humaine se heurtent à des difficultés considérables qui sont : la grande complexité du spectre de fréquence des sons émis, la proximité des spectres de phonèmes voisins, la multiplicité des différents phonèmes utilisés dans une même langue et a fortiori dans des langues et dialectes différents, et surtout la multiplicité des manières dont les sons sont effectivement émis en fonction des sons qui précèdent ou suivent (phénomènes de coarticulation).

Il est donc très difficile soit de reconnaître une succession de phonèmes émis à grande cadence, pour reconstituer les mots qui ont été prononcés, soit de synthétiser des suites de sons et des mots qui seront effectivement reconnus avec leur signification par ceux qui les entendent.

Un procédé connu pour synthétiser la parole consiste à utiliser un dispositif simulant le comportement d'un tube acoustique à section variable qui représente le conduit vocal par lequel la parole humaine est émise.

Le conduit vocal, qui part de ce qu'il est convenu d'appeler les cordes vocales (qui agissent comme source d'excitation à l'extrémité amont du tube) s'étend du larynx aux lèvres, en passant par le pharynx et la cavité buccale. C'est un conduit dont la section n'est pas uniforme sur sa longueur et varie dans de larges limites (par exemple 2 cm² dans le larynx, de 3 à 7 cm² dans le pharynx, de 0 à 15 cm² pour la cavité buccale, 0 cm² aux lèvres si les lèvres sont fermées, etc).

Ce conduit vocal peut être représenté comme un tube acoustique constitué par une succession de tronçons individuels de longueur constante dont la section au repos a une valeur déterminée.

Les livres de G. FANT, *Acoustic Theory of Speech Production*, 1960, Mouton and CO, Gravenhage, Pays Bas, et J. L. FLANAGAN, *Speech Analysis Synthesis and Perception*, 1972, SPRINGER-VERLAG - New-York, font état de ce type de représentation dans lesquelles le conduit vocal est découpé en tronçons successifs de l'ordre d'un centimètre de long dont les sections peuvent être répertoriées. La production des sons peut se modéliser par des variations des aires de sections des tronçons individuels.

On peut donc produire des sons se rapprochant des phonèmes de la parole humaine en utilisant une succession de tronçons de tubes acoustiques avec une source de débit d'air à l'entrée, cette source ayant des caractéristiques similaires à celles des cordes vocales humaines, et en faisant varier les sections des différents tronçons.

Bien entendu, avec les techniques modernes de traitement de signal par ordinateur, on n'utilisera pas un tube acoustique matériel ayant des sections pouvant matériellement varier tronçon par tronçon, mais on utilisera une simulation de la source d'air et

du conduit vocal soit par des circuits électriques analogiques, soit par un ordinateur dans lequel on pourra faire varier des paramètres représentant notamment les sections de tube, la longueur totale du tube, le spectre du débit d'air de la source.

Le ordinateur fournira en sortie vers un haut-parleur (pour la synthèse de la parole) un signal électrique dont le spectre et les variations de spectre reproduisent aussi fidèlement que possible le spectre et les variations de spectre du son ou des successions de sons qu'on veut émettre. Pour l'analyse de la parole, c'est un microphone qui reçoit le message acoustique, qui le convertit en signaux électriques, reçus et traités par le ordinateur, par exemple après des conversions analogiques/numériques pour que le ordinateur puisse travailler en mode numérique. Le résultat de l'analyse peut être utilisé directement en reconnaissance de parole ou être codé et transmis pour reconstitution. Le codage peut être de type scalaire ou vectoriel.

Si le principe de la simulation du conduit vocal par une succession de tronçons de tubes acoustiques de section variable est connu, il n'a jamais pu être mis en oeuvre de manière satisfaisante pour permettre l'analyse ou la synthèse de la parole continue. Le plus souvent quelques essais sont faits par exemple avec des voyelles ou des ensembles consonne-voyelle ; mais on est très loin de parvenir à synthétiser ou reconnaître des successions rapides de sons tels qu'ils se présentent dans la parole humaine.

La raison en est que la commande automatique à partir d'un texte est difficile et mal connue ; le tube acoustique doit être paramétré par un grand nombre de facteurs : il y a beaucoup de tronçons de tubes, chacun peut subir des variations de section dans de très larges proportions (il n'y a qu'à prononcer un [o] ou un [a] pour voir la variation de section de passage de l'air entre les lèvres), et, si on appelle fonction d'aire la courbe des valeurs d'aires de section des tronçons de tube le long de la succession de tronçons, il n'y a pas de relation directe entre les fonctions d'aire du tube acoustique et les sons émis.

Par ailleurs, les spectres des sons émis dans la parole humaine sont caractérisés par des "formants" (qui sont des maxima successifs présents dans le spectre : premier formant pour la fréquence de résonance la plus basse, deuxième formant, troisième formant,...). Ces formants représentent des résonances du conduit vocal, résonances qui modulent le spectre de la source sonore (les cordes vocales) pour aboutir à un spectre modulé en sortie de conduit vocal. Les voyelles par exemple sont caractérisées par des valeurs fixes des fréquences de formants (c'est-à-dire les valeurs des fréquences d'amplitude maximale du spectre). Les consonnes sont plutôt caractérisées par les variations relatives des fréquences de formants.

Mais la combinaison d'une suite de syllabes est difficile à formaliser sous forme de variations de fréquences des formants car pour un élément de la

suite considérée les fréquences des formants dépendent des sons précédents et suivants (phénomène de coarticulation).

Des synthétiseurs de parole dits "synthétiseurs à formants" ont pu être produits : ils consistent à utiliser (ou simuler) des circuits résonants dont la fréquence de résonance peut être commandée individuellement. En combinant plusieurs fréquences de résonances correspondant aux fréquences de formants d'une voyelle déterminée, on peut synthétiser cette voyelle. En faisant varier les fréquences de résonance des circuits de la même manière que varient les fréquences de formant d'une consonne, on peut reproduire artificiellement cette consonne.

En général la connaissance des trois premiers formants ou de leurs variations au cours du temps représente une bonne approximation pour analyser ou synthétiser les sons. Mais on pourrait se contenter de deux formants pour une analyse ou synthèse simplifiée, ou au contraire aller jusqu'à quatre formants, voire plus, pour une analyse ou synthèse plus élaborée.

Dans la synthèse à formants, on analyse ou on reconstitue des spectres de signaux présentant des maxima d'amplitude pour des fréquences déterminées, mais évidemment on ne sait pas analyser ou reconstituer exactement tout le spectre et les variations de spectre qui définissent exactement la constitution du son considéré. Et le problème se complique bien sûr considérablement si, par suite des phénomènes de coarticulation entre voyelles et consonnes successives, les spectres et variations de spectre du signal se mélangent.

La présente invention part de la remarque qu'on peut combiner d'une manière tout à fait originale de particulièrement intéressante les propositions d'analyse et de synthèse de parole utilisant la simulation par un tube acoustique à section variable et les connaissances qu'on a acquises dans l'analyse et la synthèse à formants, pour aboutir à des dispositifs d'analyse et synthèse extrêmement efficaces. Leur efficacité vient du fait qu'ils fournissent une très bonne représentation des sons tout en minimisant le nombre de paramètres de représentation de ces sons et du fait qu'ils fonctionnent sur un mode qui semble très similaire au mode de fonctionnement de la parole humaine.

Selon l'invention, on propose un appareil d'analyse, de codage ou de synthèse de parole utilisant un dispositif de simulation du comportement acoustique d'un tube constitué par une succession de tronçons de sections différentes et variables mis bout à bout, caractérisé en ce que le tube comprend un ensemble de N tronçons, divisé en sous-ensembles de rangs successifs de la manière suivante : l'ensemble de N tronçons est divisé en deux sous-ensembles de rang 1, le premier sous-ensemble, du côté amont du tube, correspondant à une sensibilité négative aux variations de section pour le premier formant et le deuxième à une sensibilité positive, chaque sous-ensemble de rang i étant divisé de la même manière en deux sous-ensembles de rang i+1 s'il y a changement de signe de la sensibilité du formant i+1 dans ce sous-ensemble,

l'un des sous-ensembles correspondant à une sensibilité négative pour le (i+1)^{ème} formant et l'autre à une sensibilité positive, chaque sous-ensemble de rang (n-1) étant enfin divisé en deux tronçons, l'un des tronçons correspondant à une sensibilité négative du n^{ème} formant et l'autre à une sensibilité positive, la sensibilité du i^{ème} formant aux variations de section d'un tronçon représentant la variation relative de la fréquence du i^{ème} formant en fonction d'une variation de section de ce tronçon ; le dispositif ayant pour paramètres de commande d'analyse ou de synthèse d'une part les variations de section de certains des tronçons de tube ainsi définis, et d'autre part la longueur totale du tube ; le dispositif recevant des signaux d'un microphone ou fournissant des signaux à un haut-parleur selon qu'il fonctionne en analyseur ou synthétiseur de parole.

Ce qui est important c'est la manière dont on subdivise le tube acoustique en tronçons successifs, manière qui est liée à l'existence de formants et à la sensibilité de ces formants aux variations de section locale du tube.

Alors que dans le passé la subdivision en tronçons était soit arbitraire soit liée à des données différentes, on propose ici une subdivision très particulière liée aux formants et fonction du nombre de formants avec lesquels l'approximation d'analyse ou de synthèse doit être faite.

Plus précisément, on montrera que si l'on veut une approximation à deux formants, c'est-à-dire une approximation analogue à celle qu'on obtient dans une analyse, codage ou synthèse à deux formants mais obtenue par simulation du comportement d'un tube à tronçons successifs de section variable, on divisera le tube en quatre tronçons de longueurs successives relatives sensiblement égales à 1/6, 1/3, 1/3, 1/6 (rapportées à une longueur unitaire de tube).

Si on veut une approximation à trois formants, on utilisera une simulation d'un tube divisé en huit tronçons de longueurs successives relatives 3/30, 2/30, 4/30, 6/30, 6/30, 4/30, 2/30, 3/30.

On montrera dans la suite comment ces valeurs sont obtenues.

Les valeurs théoriques de ces longueurs peuvent être calculées précisément, mais bien entendu les valeurs pratiques peuvent être seulement des approximations des valeurs théoriques sans changer fondamentalement le résultat global d'analyse ou synthèse de parole.

Pour la détermination de la sensibilité des formants aux variations de sections, on peut faire l'approximation suivante, consistant à tracer la fonction de sensibilité du formant aux variations de section d'un tronçon en fonction de la position de ce tronçon entre l'extrémité amont et l'extrémité aval du tube.

Pour le premier formant, cette fonction peut être assimilée à une demi-période de sinuséide, la sensibilité étant négative et maximale à l'entrée amont du tube, nulle au milieu, et positive et maximale à la sortie. Par sensibilité positive on entend une croissance de la fréquence du formant pour une croissance de la section considérée. Une sensibilité négative est une décroissance de fré-

quence pour une croissance de section.

Pour le deuxième formant, la fonction de sensibilité peut être assimilée à trois demi-périodes de sinusoïde entre l'entrée et la sortie. Pour le $i^{\text{ème}}$ formant, la fonction peut être assimilée à une sinusoïde dont la demi-période est $L/(2i-1)$ où L est la longueur totale du tube, la sensibilité étant maximale et négative à l'entrée amont (il y a donc $2i-1$ demi-périodes entre l'entrée et la sortie du tube pour la fonction de sensibilité du $i^{\text{ème}}$ formant).

Les zones de passage vers zéro de la sensibilité des différents formants constituent les frontières des tronçons de tubes successifs. Il y a $N = 2 + n(n-1)$ tronçons au total si on fait une approximation à n formants.

L'action sur les sections des tronçons de tube du dispositif de simulation peut s'exercer de plusieurs manières différentes :

- action sur la section globale du tronçon
- action sur la section d'une portion locale de tronçon située vers le milieu du tronçon (pour agir sur tous les formants à la fois)
- action sur la section d'une portion locale de tronçon située à la frontière entre deux tronçons (si on veut supprimer volontairement l'action sur l'un des formants : celui dont la sensibilité s'annule à cette frontière).

Par cette organisation judicieuse de tronçons de tubes bien choisis, on a lié directement l'analyse et la synthèse de la parole humaine à la notion de formants, ce qui permet de minimiser le nombre de paramètres de commande du dispositif de simulation lorsqu'on veut produire des sons dont on a justement répertorié les formants et leurs variations.

Cette organisation se distingue donc fondamentalement des propositions déjà faites en matière de simulation par tubes de section variable puisque jusqu'à maintenant on se contentait de subdiviser les tubes en tronçons d'une manière artificielle : subdivision en tronçons réguliers de l'ordre de 1 cm de long ou, par analogie avec le conduit vocal, subdivision entre une zone de larynx, de pharynx et subdivision arbitraire dans la bouche.

D'autres caractéristiques et avantages de l'invention apparaîtront à la lecture de la description détaillée qui suit et qui est faite en référence aux dessins annexés dans lesquels :

- la figure 1 représente la forme générale d'un conduit vocal humain ;
- la figure 2 représente la schématisation de ce conduit sous forme d'un tube divisé en tronçons de sections différentes, variables individuellement ;
- la figure 3 représente le schéma-bloc d'un dispositif de synthèse de parole ;
- la figure 4 représente le tracé des courbes de sensibilité des quatre premiers formants d'un tube uniforme ;
- la figure 5 représente la division d'un tube selon l'invention en quatre tronçons pour une approximation limitée aux deux premiers formants ;
- la figure 6 représente la division d'un tube selon l'invention en huit tronçons pour une approximation limitée aux trois premiers for-

mants ; et

- la figure 7 représente la division d'un tube selon l'invention en quatorze tronçons pour une approximation limitée aux quatre premiers formants.

la figure 1 représente en coupe l'anatomie simplifiée d'un conduit vocal humain avec différentes régions et organes tels que les cordes vocales CV constituant la source de débit d'air (de forme d'onde périodique très particulière), la luette LU, le palais PL, la langue LN, les dents DN, les lèvres supérieure LS et inférieure LI.

La figure 2 représente une schématisation qu'on a pu faire du conduit vocal sous forme d'un tube acoustique 10 composé de tronçons cylindriques T1, T2...T16, adjacents, ayant des sections différentes les unes des autres au repos, ces sections pouvant varier indépendamment les unes des autres ; la combinaison des variations de section des différents tronçons permet de produire des sons. Les voyelles se traduisent essentiellement par des rapports entre les différentes sections. Les consonnes se traduisent plutôt par des transitions entre une première combinaison de sections et une deuxième combinaison.

Pour la synthèse de parole, le tube est placé derrière une source de débit d'air reproduisant les caractéristiques des cordes vocales, c'est-à-dire notamment une onde de débit périodique de période environ 10 millisecondes ayant une forme de dents de scie très arrondies, le front de croissance étant plus lent que le front de décroissance.

Compte-tenu de la difficulté de réaliser mécaniquement un tel tube acoustique, on utilisera plutôt des technologies modernes de simulation par ordinateur, dans lesquelles le comportement acoustique du tube peut être déterminé c'est-à-dire que le débit et la pression d'air en chaque point et notamment à la sortie du tube, peuvent être calculés ; les caractéristiques du signal électrique qu'il faut appliquer à un haut-parleur pour reproduire ce débit et cette pression sont également calculées, et un signal électrique ayant ces caractéristiques est produit par un générateur commandé par le calculateur.

La figure 3 représente très schématiquement ce mode de réalisation matérielle d'un synthétiseur de parole par simulation : un organe d'entrée de données définit la succession des phonèmes à produire. Cet organe peut être par exemple un clavier alphanumérique CL sur lequel des touches ou des combinaisons de touches représentent des phonèmes.

Ces données sont appliquées au calculateur CALC sous forme de signaux électriques de manière classique, par l'intermédiaire d'un bus de liaison.

Le calculateur contrôle un synthétiseur de signaux électriques (GEN) qui lui-même commande un haut-parleur HP.

Le fonctionnement du calculateur est contrôlé de la manière suivante : à partir de la suite des commandes du clavier, un ensemble de paramètres est engendré ; ces paramètres correspondent aux valeurs des sections des tronçons du tube acoustique représentant le conduit vocal et aux variations

de ces sections au cours du temps.

Le traitement de données est simplement la simulation par calcul du comportement du tube ayant ces sections et ces variations de sections. Ce comportement est maintenant bien connu et il est décrit par exemple dans l'ouvrage de J.L. Flanagan susmentionné.

Le traitement aboutit d'abord à des résultats de débit d'air et/ou pression d'air en sortie du tube puis à la génération des caractéristiques du signal électrique qu'il faut appliquer à un haut-parleur pour reproduire la pression en sortie. On peut supposer pour simplifier que la pression d'air provoquée par le haut-parleur est proportionnelle au courant électrique instantané qui l'alimente. Dans ce cas, le traitement consiste à déterminer à chaque instant quelle est la forme d'onde de la pression d'air représentant le son désiré, le synthétiseur de signaux électriques fournissant une forme d'onde de courant correspondant exactement à la forme d'onde de la pression d'air calculée. Bien entendu, si le haut-parleur a une courbe de réponse pression d'air/courant électrique non linéaire, le calcul doit en tenir compte.

Etant donné que l'invention ne porte pas sur le principe de synthèse ou d'analyse de la parole par simulation du comportement acoustique d'un tube, principe qui est connu, mais sur le choix des paramètres de la simulation, on va maintenant détailler ce choix.

Le choix porte sur les longueurs des tronçons de tubes utilisés dans le traitement de données.

C'est-à-dire que les paramètres stockés en mémoire dans le calculateur ne seront pas les variations de section de tronçons d'un tube découpé arbitrairement en tronçons de longueurs quelconques (comme c'est le cas sur la figure 2 où on a pris par commodité des tronçons qui ont tous la même longueur) mais ces paramètres représenteront les variations de section de tronçons de longueurs bien déterminées résultant du découpage selon l'invention qu'on va maintenant expliquer en détail.

On part d'un tube de longueur totale L (par exemple de 15 à 20 cm, ce qui correspond à la longueur du conduit vocal). La réponse acoustique de ce tube présente des formants, c'est-à-dire des résonances plus ou moins prononcées à certaines fréquences. Le spectre d'un signal acoustique émis à l'entrée du tube sera modulé par ces formants et présentera des maxima locaux aux fréquences des formants.

L'étude acoustique théorique d'un tube de longueur L montre que la fréquence des formants varie en fonction de la section du tube. Mais elle ne varie pas de la même manière partout : si on fait varier la section du tube uniquement localement au milieu de la longueur du tube, on s'aperçoit que la fréquence des formants ne varie pas du tout ; si, au contraire, on fait varier la section uniquement à l'embouchure du tube ou à sa sortie, on constate qu'une variation de section fait varier la fréquence des formants : si c'est à l'embouchure du tube que la section varie, la fréquence des formants augmente à mesure que la section diminue ; si, au contraire, c'est à la sortie du

tube que la section varie, la fréquence des formants augmente à mesure que la section augmente.

Enfin, si on fait varier la section du tube en un endroit quelconque, les fréquences des différents formants varieront avec des amplitudes et des sens différents.

En fait, pour un tube initialement à section uniforme, on peut donner une représentation théorique de la sensibilité des formants, c'est-à-dire du sens de variation des fréquences de formant en fonction d'une variation locale de section du tube, car la sensibilité des formants varie sinusoidalement le long du tube entre l'embouchure et la sortie, la période de la sinusoïde étant différente pour chacun des formants.

C'est ce qui est représenté à la figure 4 : le diagramme 4a représente la courbe de sensibilité SF1 du premier formant F1 du tube en fonction de la position x (x variant entre 0 et L) à laquelle on produit une variation de section.

Le diagramme 4b représente la courbe de sensibilité SF2 du deuxième formant F2, le diagramme 4c représente la courbe de sensibilité SF3 du troisième formant F3, et le diagramme 4d représente la courbe de sensibilité SF4 du quatrième formant F4.

Sur ces courbes, on ne s'est pas préoccupé de la valeur relative des sensibilités SF1, SF2, SF3, SF4 les unes par rapport aux autres. Seuls la forme de variation, les signes, les positions des maxima et minima et des passages par zéro nous intéressent selon l'invention. On a donc donné une valeur maximale unitaire à chacune des sensibilités.

La forme théorique des courbes de sensibilité des formants en fonction de la position x à laquelle une variation de section est appliquée est très simple : c'est une sinusoïde dont la demi-période est $L/(2i-1)$ où i est le rang du formant : $i = 1$ pour le premier formant F1, c'est-à-dire pour la fréquence de résonance la plus basse ; $i = 2$ pour la fréquence de résonance suivant immédiatement ; et ainsi de suite. Cette sinusoïde présente un minimum (sensibilité maximale en valeur absolue mais négative) à l'embouchure du tube ($x = 0$) et un maximum (sensibilité maximale et positive) à l'extrémité du tube ($x = L$).

On peut vérifier que le tube est antisymétrique, c'est-à-dire qu'une action sur la section en un point quelconque d'abscisse x agit sur les différents formants exactement de la même manière, mais avec un signe opposé, qu'une action sur la section en un point d'abscisse L-x.

Pour $x = L/2$ l'action est donc nulle : la sensibilité passe par zéro en ce point pour tous les formants quel que soit leur rang.

Cette remarque sera importante pour la suite car elle permettra de limiter le nombre de paramètres de commande du dispositif d'analyse ou de synthèse de parole : on obtient la même variation de fréquences de formants, pour tous les formants à la fois en agissant sur la section au point d'abscisse x au lieu du point d'abscisse L-x à condition de faire varier la section en ce point dans le sens opposé à celui qu'on aurait utilisé au point L-x.

Les explications ci-dessus ont été dans le cadre

d'un tube initialement à section uniforme aux tronçons duquel on applique de petites variations. Des expériences effectuées par les inventeurs ont montré que dans le cas d'un tube divisé en tronçons de sections variables et dans le cas où des variations importantes sont appliquées à ces sections, les sens de variations sont conservés même si les fonctions de sensibilité ne sont plus sinusoidales.

L'invention propose de diviser le tube en tronçons dont les limites correspondent exactement aux passages par zéro de la sensibilité des formants avec lesquels on veut faire un approximation d'analyse ou de synthèse de parole : chaque passage par zéro définit la limite d'un tronçon.

Les passages par zéro de la sensibilité des formants sont situés aux abscisses :

- A0 pour le premier formant F1
- B1, A0, B'1 pour le deuxième formant F2
- C1, C2, A0, C'2, C'1 pour le troisième formant F3
- D1, D2, D3, A0, D'3, D'2, D'1 pour le quatrième formant F4 et ainsi de suite.

Les valeurs de ces abscisses sont les suivantes :

A0 = L/2	(milieu du tube)
B1 = L/6	B'1 = L - L/6
C1 = L/10	C'1 = L - L/10
C2 = 3L/10	C'2 = L - 3L/10
D1 = L/14	D'1 = L - L/14
D2 = 3L/14	D'2 = L - 3L/14
D3 = 5L/14	D'3 = L - 5L/14

On va donner trois exemples de découpage selon l'invention puis une règle générale :

Premier exemple : on veut une approximation à deux formants F1 et F2.

Le tube est découpé en quatre tronçons qui sont :

- un premier tronçon de 0 à B1 (longueur L/6)
- un deuxième tronçon de B1 à A0 (longueur L/3)
- un troisième tronçon de A0 à B'1 (longueur L/3)
- un quatrième tronçon de B'1 à L (longueur L/6)

Le tube correspondant est représenté à la figure 5.

Deuxième exemple : on veut une approximation à trois formants F1, F2, F3.

Le tube est divisé en huit tronçons qui sont :

- un premier tronçon de 0 à C1 (longueur L/10)
- un deuxième tronçon de C1 à B1 (longueur L/15)
- un troisième tronçon de B1 à C2 (longueur 2L/15)
- un quatrième tronçon de C2 à A0 (longueur 3L/15)
- et quatre autres tronçons symétriques des quatre premiers par rapport au milieu du tube.

Le tube est représenté à la figure 6.

Troisième exemple : on veut une approximation à quatre formants F1, F2, F3, F4.

Le tube est divisé en 14 tronçons qui sont représentés à la figure 7 et qui sont :

- un premier tronçon de 0 à D1 (longueur L/14)
- un deuxième tronçon D1 à C1 (longueur L/35)
- un troisième tronçon C1 à B1 (longueur L/15)

- un quatrième tronçon de B1 à D2 (longueur L/21)
- un cinquième tronçon D2 à C2 (longueur 3L/35)
- un sixième tronçon C2 à D3 (longueur 2L/35)
- un septième tronçon D3 à A0 (longueur L/7)
- et sept autres tronçons symétriques des premiers par rapport au milieu du tube.

Pour généraliser la méthode à une approximation à n formants (bien qu'il soit peu probable qu'on veuille dépasser n = 4), on détermine l'abscisse $X_{i,j}$ du $j^{\text{ème}}$ passage par zéro de la sensibilité du $i^{\text{ème}}$ formant, pour tous les formants ($i = 1$ à n) et sur toute la longueur du tube ($j = 1$ à $2i - 1$).

On a $X_{i,j} = L(2j - 1) / (2i - 1) \times 2$.

On classe tous les $X_{i,j}$ dans l'ordre croissant le long du tube à leurs positions respectives ; chaque tronçon de tube est délimité par deux abscisses adjacentes de cette suite classée, le premier tronçon commençant à l'abscisse 0 et se terminant à l'abscisse $X_{n,1} = L/2n-1$ et le dernier tronçon commençant à l'abscisse $X_{n,2n-1} = L - L/(2n-1)$ et se terminant à l'abscisse L.

Le nombre total de tronçons est $N = n(n-1) + 2$.

On a ainsi défini précisément une série de paramètres très importants pour faire fonctionner le dispositif d'analyse ou synthèse de parole, ces paramètres étant le nombre de tronçons et la longueur de chacun.

Ces paramètres sont fournis au calculateur et le traitement de données consiste en une action sur la section des tronçons définis par ces paramètres. L'action peut porter sur un nombre de tronçons égal à la moitié du nombre total, pour la raison de symétrie indiquée précédemment.

Par des études détaillées on déterminera quelles sont les variations de section à effectuer sur chaque tronçon pour produire tel ou tel phonème (et on est guidé pour cela par les connaissances déjà établies sur les fréquences de formants et variations de fréquences de formants correspondant à ces phonèmes).

Une mémoire de données pourra être associée au calculateur, mémoire contenant directement pour chaque phonème les séquences de variations de section des tronçons ainsi définis.

Dans un dispositif de synthèse de parole, le déclenchement de ces séquences de variation aboutit, après traitement dans la calculateur, à la génération de signaux électriques transmis au haut-parleur, et à la production du phonème désiré.

Dans un dispositif d'analyse de parole, on procède par rebouclage : un microphone reçoit les sons, les convertit en signaux électriques. Ces signaux sont traités par le calculateur. Une comparaison est effectuée entre des données issues du traitement et des données générées par des séquences de variations de sections correspondant à des sons connus.

L'invention peut être utilisée comme jouet éducatif de synthèse de parole permettant de mieux comprendre l'élaboration des sons par le système vocal humain. Dans ce cas, la source pourra être une embouchure comprenant une anche dans laquelle l'utilisateur soufflera. On pourra aussi utiliser une source de bruit blanc. On utilisera 4 ou 8 tronçons dont les volumes sont contrôlés par des pistons

commandés par les doigts de la main. L'appareil pourra être fabriqué par moulage plastique.

Revendications

1. Appareil d'analyse, de codage ou de synthèse de parole utilisant un dispositif de simulation du comportement acoustique d'un tube constitué par une succession de tronçons (T1, T2...) de sections différentes et variables mis bout à bout, caractérisé en ce que le tube comprend un ensemble de N tronçons, divisé en sous-ensembles de rangs successifs de la manière suivante : l'ensemble de N tronçons est divisé en deux sous-ensembles de rang 1, le premier sous-ensemble, du côté amont du tube, correspondant à une sensibilité négative aux variations de section pour le premier formant et le deuxième à une sensibilité positive, chaque sous-ensemble de rang i étant divisé de la même manière en deux sous-ensembles de rang i+1 s'il y a changement de signe de la sensibilité du formant i+1 dans ce sous-ensemble, l'un des sous-ensembles correspondant à une sensibilité négative pour le (i+1)^{ème} formant et l'autre à une sensibilité positive, chaque sous-ensemble de rang (n-1) étant enfin divisé en deux tronçons, l'un des tronçons correspondant à une sensibilité négative du n^{ème} formant et l'autre à une sensibilité positive, la sensibilité du i^{ème} formant aux variations de section d'un tronçon représentant la variation relative de la fréquence du i^{ème} formant en fonction d'une variation de section de ce tronçon ; le dispositif ayant pour paramètres de commande d'analyse ou de synthèse, d'une part, les variations de section de certains des tronçons de tube ainsi définis, et, d'autre part, la longueur totale L du tube ; le dispositif recevant des signaux d'un microphone ou fournissant des signaux à un haut-parleur selon qu'il fonctionne en analyseur ou synthétiseur de parole.

2. Appareil d'analyse ou synthèse de la parole selon la revendication 1, caractérisé en ce que le nombre de tronçons du tube est $N = 2 + n(n-1)$ si on veut faire une synthèse ou une analyse de parole avec une approximation correspondant aux n premiers formants du tube.

3. Appareil d'analyse ou synthèse de parole selon la revendication 2, caractérisé en ce que $n = 2$, et en ce que le tube est divisé en $N = 4$ tronçons dont les longueurs successives sont L/6, 2L/6, 2L/6 et L/6 respectivement.

4. Appareil d'analyse ou synthèse de parole selon la revendication 2, caractérisé en ce que $n = 3$, et en ce que le tube est divisé en $N = 8$ tronçons dont les longueurs successives sont L/10, L/15, 2L/15, 3L/15, 3L/15, 2L/15, L/15 et L/10 respectivement.

5. Appareil d'analyse ou synthèse de parole selon la revendication 2, caractérisé en ce que $n = 4$, et en ce que le tube est divisé en $N = 14$ tronçons dont les longueurs successives sont

L/14, L/35, L/15, L/21, 3L/35, 2L/35, L/7, L/7, 2L/35, 3L/35, L/21, L/15, L/35 et L/14.

6. Appareil de synthèse de parole selon la revendication 1, caractérisé en ce qu'il est réalisé matériellement sous forme d'une succession de cavités correspondant auxdits tronçons, des moyens de réglage et de commande du volume de chaque cavité étant prévue.

7. Appareil de synthèse de parole selon la revendication 6, caractérisé en ce que la cavité amont est associée à une embouchure dans laquelle un utilisateur peut souffler et en ce que les moyens de réglage comprennent des pistons actionnables manuellement.

5

10

15

20

25

30

35

40

45

50

55

60

65

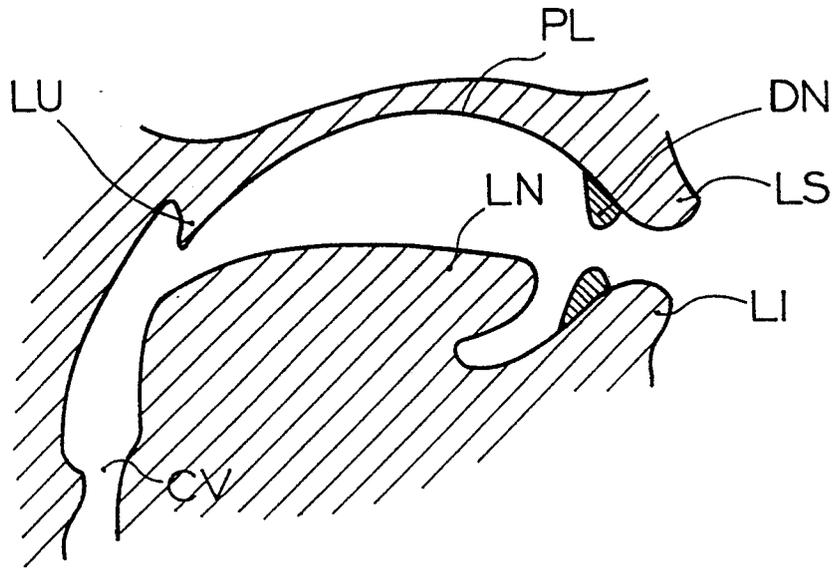


Fig. 1

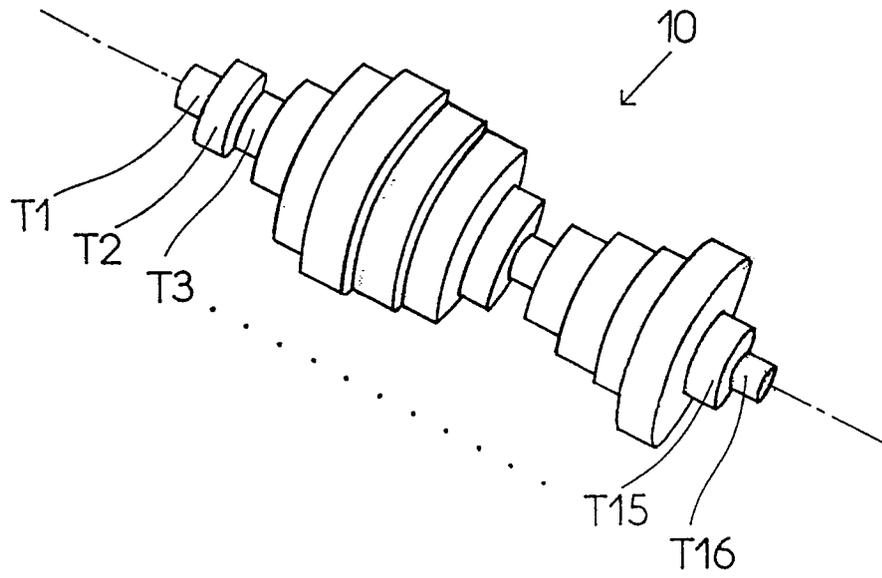


Fig. 2

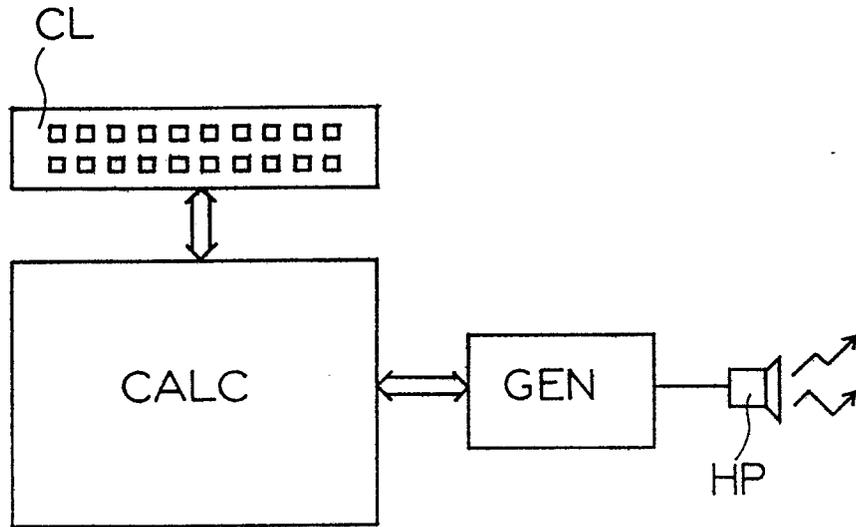


Fig. 3

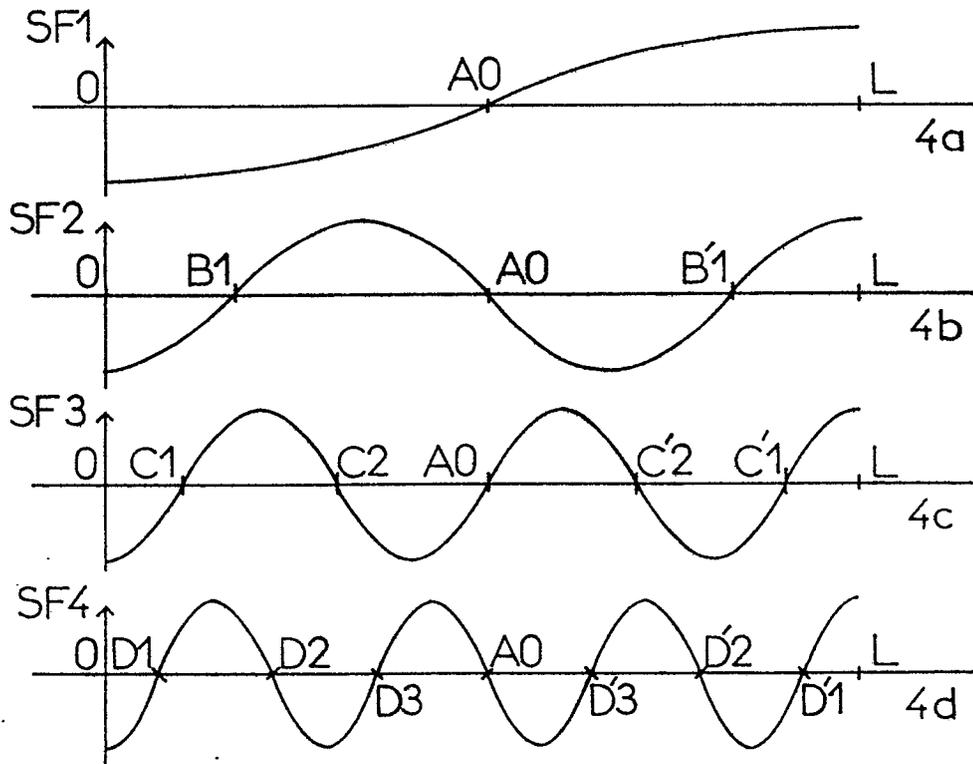


Fig. 4

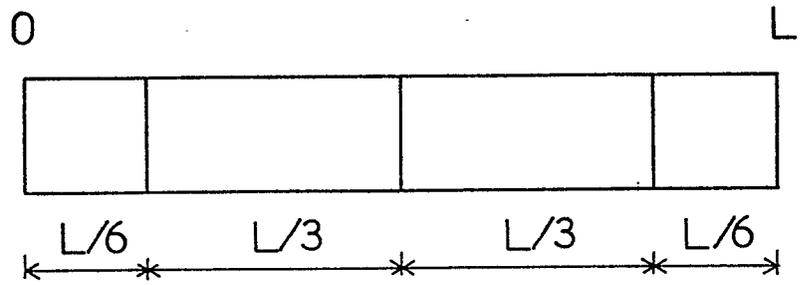


Fig.5

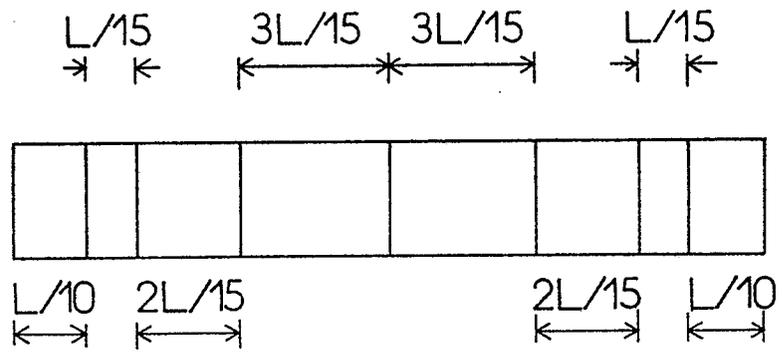


Fig.6

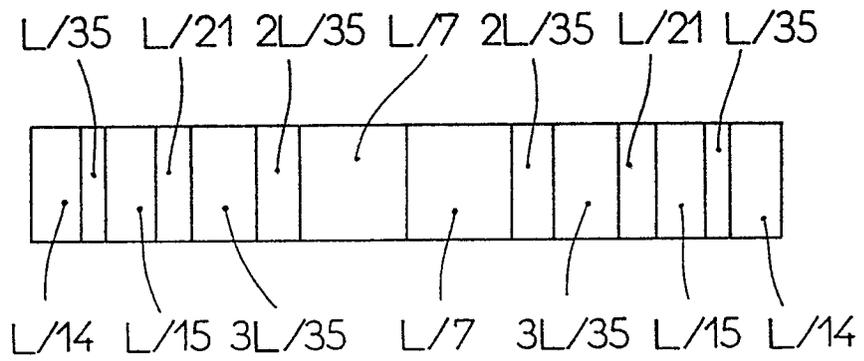


Fig.7