



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 0 953 969 B1**

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention
de la délivrance du brevet:
01.10.2003 Bulletin 2003/40

(51) Int Cl.7: **G10L 19/00, G10L 21/04**

(21) Numéro de dépôt: **99400873.8**

(22) Date de dépôt: **09.04.1999**

(54) **Procédé de restitution de la parole à régulation de silences**

Verfahren zur Sprachwiedergabe mit Regelung von Pausen

Method for rendering speech with silence regulation

(84) Etats contractants désignés:
DE GB

(30) Priorité: **27.04.1998 FR 9805227**

(43) Date de publication de la demande:
03.11.1999 Bulletin 1999/44

(73) Titulaire: **SAGEM S.A.**
75116 Paris (FR)

(72) Inventeur: **Charbonnier, Philippe**
78600 Le Mesnil le Roi (FR)

(74) Mandataire: **Bloch, Gérard**
2, square de l'Avenue du Bois
75116 Paris (FR)

(56) Documents cités:
EP-A- 0 756 267 US-A- 5 526 353
US-A- 5 682 384

- **RAMACHANDRAN RAMJEE ET AL: "ADAPTIVE
PLAYOUT MECHANISMS FOR PACKETIZED
AUDIO APPLICATIONS IN WIDE-AREA
NETWORKS" PROCEEDINGS OF THE
CONFERENCE ON COMPUTER
COMMUNICATIONS (INFOCOM), TORONTO,
JUNE 12 - 16, 1994, vol. 2, 12 juin 1994
(1994-06-12), pages 680-688, XP000496524
INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS**
- **DEMPSEY ET AL.: "Destination Buffering for
Low-Bandwidth Audio Transmission using
Redundancy-Based Error Control"
PROCEEDINGS OF THE 21ST IEEE
CONFERENCE ON LOCAL COMPUTER
NETWORKS, 13 - 16 octobre 1996, pages
345-354, XP002088688 MINNEAPOLIS**

Il est rappelé que: Dans un délai de neuf mois à compter de la date de publication de la mention de la délivrance du brevet européen, toute personne peut faire opposition au brevet européen délivré, auprès de l'Office européen des brevets. L'opposition doit être formée par écrit et motivée. Elle n'est réputée formée qu'après paiement de la taxe d'opposition. (Art. 99(1) Convention sur le brevet européen).

EP 0 953 969 B1

Description

[0001] Dans le réseau téléphonique commuté, la parole est transmise sous forme d'un flux continu de mots de code représentant l'amplitude instantanée du signal vocal, échantillonné cycliquement et numérisé à cet effet. Une communication dispose en permanence d'une voie de transmission, ou circuit, pour écouler ce flux continu.

[0002] Par contre, sur un réseau informatique du type de l'Internet, ou sur certains réseaux radio en mode paquet, les mots de code représentant la voix sont transmis par paquets, sur un canal offrant un débit suffisant pour pouvoir partager temporellement le canal entre plusieurs terminaux.

[0003] La transmission par paquets à travers un réseau engendre cependant des problèmes.

[0004] En effet, le temps de traversée du réseau est essentiellement variable, car chaque paquet peut suivre un chemin variable et on ne peut pas prendre en compte les paquets dès leur réception, car un paquet peut arriver avant la fin de la restitution sonore du précédent, ou peut, au contraire, arriver après celle-ci. Dans le premier cas, il y aurait, si prise en compte, écrasement de paquet, dans le deuxième cas, temps mort.

[0005] On interpose donc une mémoire tampon dans laquelle on stocke les paquets au rythme aléatoire de leurs arrivées et on lit les paquets, pour la restitution sonore, au rythme fixe de l'émetteur de ceux-ci, avec un retard, par rapport à leurs instants d'émission, qui est fixe et suffisamment grand, par rapport aux fluctuations des durées de transmission, pour que les paquets aient été reçus.

[0006] Il peut même se produire que les paquets soient reçus dans un ordre différent de celui de leur émission, si bien qu'il faut les numérotter à l'émission pour rétablir l'ordre voulu en restitution.

[0007] Tout cela nécessite que la mémoire tampon soit de taille relativement grande. Cependant ce retard de la restitution sonore devient perceptible et gênant pour le dialogue entre les deux interlocuteurs.

[0008] Un exemple de restitution avec correction est connu de US-A-5 526 353

[0009] La présente invention vise à remédier à ces inconvénients.

[0010] A cet effet, l'invention concerne un procédé de restitution sonore de signaux de parole, reçus par paquets successifs représentant des tranches temporelles successives de parole comportant à l'origine des périodes de silence et mémorisés temporairement avant d'être restitués de façon sonore, procédé caractérisé par le fait qu'on détecte la présence desdits silences, dans les tranches reçues, et qu'on en régule la durée de restitution pour restituer, d'un seul tenant, les signaux de parole autres que les silences.

[0011] Ainsi, on module le temps, au niveau de la restitution, pour, en pratique, compenser la fluctuation sur les instants d'arrivée des paquets. On réunit ainsi tem-

porellement les parties d'une séquence sonore d'origine qui ont été séparées physiquement, et donc temporellement, de par leur mise dans des paquets différents. Comme la modulation du temps porte sur les silences, elle est en pratique sans inconvénient pour la compréhension des paroles.

[0012] Avantageusement, on abrège la restitution de tout silence qui est suivi d'une séquence sonore complète à restituer.

[0013] L'invention sera mieux comprise à l'aide de la description suivante d'une forme de réalisation préférée du procédé de l'invention, en référence au dessin annexé, sur lequel :

- 15 - la figure 1 est un schéma par blocs illustrant un dispositif de mise en oeuvre du procédé de l'invention,
- la figure 2, formée des figures 2A, 2B, 2C, 2D et 2E, illustre le découpage par paquets d'un signal de parole en fonction du temps t , et
- 20 - la figure 3 est un diagramme de cheminement illustrant le procédé.

[0014] Le dispositif de la figure 1 comporte un automate 1 à microprocesseur commandant le fonctionnement d'une chaîne de réception et de restitution sonore de signaux de parole. La chaîne, reliée en entrée à une ligne 2 d'un réseau de communication par paquets, comporte en entrée un circuit 3 de lecture des paquets reçus, qui repère les segments sonores et les silences et reconstitue leur position temporelle. Selon le type de codage employé pour représenter la voix, des repères explicites peuvent exister qui distinguent les segments et indiquent leurs dates, ou bien de tels repères explicites n'existent pas et le circuit 3 les reconstitue à partir du numéro d'ordre des paquets et de leur contenu décodé en signal vocal dont il analyse la forme. Le circuit 3 est suivi d'une mémoire tampon 4 de stockage des paquets, qui remet ceux-ci dans leur ordre d'émission et les transmet à un circuit 5 de restitution ou reproduction sonore commandant un écouteur 6.

[0015] Le fonctionnement du dispositif ci-dessus va maintenant être expliqué.

[0016] De façon classique, un terminal émetteur, en communication avec celui de la figure 1, analyse le signal de son microphone par un vocodeur pour le coder sous forme comprimée.

[0017] La figure 2A représente l'amplitude S du signal vocal en fonction du temps t .

[0018] Dans une variante explicite, un vocodeur cherche à délimiter des segments comportant un signal vocal correspondant à un son normalisé en bibliothèque, comme par exemple une voyelle, une consonne ou une quasi-silence. Pour la transmission d'informations représentant le signal S , celui-ci est alors remplacé par une suite de mots de code représentant les sons qui y ont été reconnus. Le volume de données d'information est ainsi très réduit. En réception, la consultation d'une bibliothèque semblable permet la restitution du signal

S d'origine.

[0019] Dans une variante implicite, le signal n'est pas analysé si finement pour le codage, et c'est le circuit 3 du récepteur qui analyse le signal reconstitué pour repérer les segments de silence. Dans tous les cas, l'information codée est transmise en paquets P1, P2, P3, P4 véhiculant chacun une tranche plus ou moins longue du discours.

[0020] Dans l'exemple de la figure 2A, le signal S comporte, dans le paquet P1, la fin S0 d'un silence, un bloc de signal énergétique de parole 11 suivi d'un silence S1 et d'un autre bloc 12. Dans le paquet P2, le bloc 12 se poursuit, avec la référence 13, et est suivi de deux blocs 14 et 15 avec des silences S2 et S3 interposés. Le paquet P3 comporte la fin du bloc 15, référencée 16, un silence S4, un bloc 17, un silence S5 et le début 18 d'un bloc suivi, paquet P4, d'une fin 19 du bloc puis du début S6 d'un silence.

[0021] Les blocs 11, 12-13, 14, 15-16, 17, 18 représentent ici les six séquences respectives : "et" "l'invention" "est" "nouvelle" "et" "inventive" (figure 2B).

[0022] En réception, le bloc 12-13 (tout comme 15-16 et 18-19), réparti sur les deux paquets, P1 et P2, risque d'être temporellement séparé en deux lors de sa restitution sonore. On cherche donc ici à l'éviter.

[0023] Les instants t0, t1, t2, t3 (fig. 2A) délimitent les tranches du signal initial affectées aux paquets successifs P1, P2, P3, P4. Les références t'0, t'1, t'2, t'3 (fig. 2C), uniformément translatées par rapport aux instants respectifs t0, t1, t2, t3, marquent les dates théoriques correspondantes de restitution dans l'écouteur 6. En raison de la fluctuation du délai de transmission, des paquets peuvent arriver en avance ou en retard. Dans cet exemple, le paquet P2 arrive après un temps mort ou retard R suivant l'instant t'1 de fin de restitution du paquet P1. Au contraire, le paquet P3, bien qu'arrivant après t'2, arrive en avance sur la fin de la restitution du paquet P2.

[0024] La figure 2D illustre une restitution qui serait immédiate. La phrase

"et... l'invention... est... nouvelle... et... inventive",

dans laquelle "." représente un silence naturel comme S1, devient :

"et...l'inven□□□tion...est...nouv/elle...et... inventive"

où "□□□" représente un silence parasite qui s'interpose (de durée quelconque), coupant le mot "invention".

et "/" représente une superposition additive, ou un écrasement, entre le début et la fin (qui arrive trop tôt) du mot "nouvelle", le mot "inventive" étant de même déformé.

[0025] La figure 2E illustre le procédé de l'invention. Sur celle-ci,

"." représente un silence d'origine qui a été abrégé et
représente un silence d'origine qui qui été été allongé.

[0026] On remarque tout d'abord que le silence parasite □□□ de la figure 2D a disparu, de même que la superposition "/".

[0027] A l'arrivée du paquet P1, le bloc sonore entier 11 ("et") est ici restitué immédiatement (on suppose qu'on était dans une phase de restitution du début du silence S0 transmis dans le paquet qui précède). Par contre, la séquence 12 ("l'inven") du bloc 12-13 n'est pas restituée à ce moment, car elle ne constitue qu'une portion de séquence. Le silence S1 est allongé (....) jusqu'à réception du paquet P2, pour lequel les séquences complètes 12-13 et 14, avec le silence S2, sont restituées. Lorsqu'arrive le paquet P3, en avance relative par rapport au paquet P2, le silence S3 est raccourci (.) pour restituer, sans délai notable, la séquence : "nouvelle.et.". Les silences S2 et S4 sont ici raccourcis, ou abrégés (.) pour vider au plus vite la mémoire tampon 4 afin de mieux tolérer des arrivées anticipées de paquets suivants. Ceci présente un intérêt surtout dans le cas pour lequel la durée de restitution de séquence(s) sonore(s) déclenchée par l'arrivée d'un paquet est supérieure à la période théorique. En effet un paquet arrivant peut compléter un bloc sonore antérieur représentant une durée ininterrompue de signal s'étendant sur plusieurs tranches de temps, qui va alors être restitué.

[0028] La figure 3 illustre la gestion à cet effet de la restitution sonore des paquets reçus.

[0029] On part ici d'un état 21 dans lequel le circuit 5 restitue une séquence sonore comme 12 - 13 ou 14. Lorsqu'arrive la fin de celle-ci, donc le début d'un silence détecté par le circuit 3, on teste, à une étape 22, la présence dans la mémoire tampon 4 de la description continue du signal jusqu'à au moins le début du segment ou bloc sonore suivant. Dans l'affirmative, on teste, à une étape 23, si le retard dépasse un seuil haut A. On appelle ici retard la durée totale de signal à restituer se trouvant en tampon; le taux de remplissage du tampon peut en constituer une approximation commode. On peut choisir un seuil A nul, c'est-à-dire passer alors de l'étape 22 à une étape 24, indiquée plus loin, de retour à l'état 21.

[0030] Dans l'affirmative à l'étape 23, on abrège, à l'étape 24, le silence en cours -ou même le supprime et restitue la séquence sonore ci-dessus par passage à l'état 21. Dans la négative, aux étapes 22 et 23, on passe à un état 26 de reproduction de silence.

[0031] On peut aussi passer à l'état 26, depuis l'état 21, en insérant un silence supplémentaire, si l'on détecte, étape 25, un débordement négatif de la mémoire 4 (absence d'information disponible en mémoire, relative

au segment suivant à reproduire).

[0032] A l'état 26, on détecte également si la mémoire 4 est vide du segment suivant à reproduire et on décide en pareil cas, étape 27, de prolonger le silence au-delà de sa durée normale, c'est-à-dire qu'on insère un silence supplémentaire.

[0033] Au contraire, étape 28, si une séquence sonore jointive, comme 12 - 13, est présente et donc disponible (paquets P1 et P2 reçus) et représente une durée dépassant un seuil B, le silence en cours à l'état 26 (S1) va être abrégé. Dans cet exemple, on fait de même si un bloc (12) de séquence sonore présente un retard à la restitution dépassant le seuil A, même en l'absence de réception de la fin (13) du bloc. Si, par contre, le seuil B n'est pas atteint, on insère un silence supplémentaire.

[0034] On quitte l'état 26 lorsque tout le silence qui devait être reproduit l'a été et qu'on doit reproduire la séquence sonore suivante. On la retarde cependant en passant transitoirement ou durablement par un état 29, d'émission d'un bruit de fond, ou d'ambiance, à consonance ou tonalité vocale du genre "euhh", traduisant que le locuteur va à nouveau parler, ce qui évite qu'on lui coupe la parole en remplaçant un silence prolongé, ou supplémentaire, par le bruit de fond. Les deux conditions de l'étape 28 (seuils B et A) sont aussi recherchées à l'état 29 et on passe, dans l'affirmative, à l'état 21 de restitution sonore.

[0035] On y teste (étape 20) si le retard à la restitution excède un seuil C, par exemple de 1,5 fois la valeur du seuil. A, ce qui indique un débordement positif de la mémoire 4. En pareil cas, les silences les plus anciens, à restituer les premiers, sont réduits et éventuellement quasiment supprimés. Il peut aussi être prévu de supprimer les séquences sonores les plus anciennes ou simplement des tranches temporelles de celles-ci, ce qui revient à moduler, ici accélérer, la vitesse de restitution du signal sonore.

Revendications

1. Procédé de restitution sonore de signaux de parole (11, S1, 12, 13, S2) reçus par paquets successifs (P1, P2, P3) représentant des tranches temporelles successives de parole comportant à l'origine des périodes de silence et mémorisés temporairement (4) avant d'être restitués de façon sonore (5, 6), procédé **caractérisé par le fait qu'on détecte la présence desdits silences (S1, S2), dans les tranches reçues, et qu'on en régule la durée de restitution pour restituer, d'un seul tenant, les signaux de parole (11, 12) autres que les silences.**
2. Procédé selon la revendication 1, dans lequel on abrège (24, 28) la restitution de tout silence (S1) dont au moins le début de la séquence sonore suivante est disponible.

3. Procédé selon la revendication 1, dans lequel on abrège (23, 28) la restitution du silence (S1) si le retard à la restitution des signaux de parole dépasse un seuil (A).

4. Procédé selon l'une des revendications 1 à 3, dans lequel on insère un silence supplémentaire (25, 27) lorsqu'aucune séquence (11) suivante n'est disponible en mémoire.

5. Procédé selon l'une des revendications 1 à 4, dans lequel, lorsqu'un silence (S1) n'est pas suivi en mémoire d'une séquence sonore de durée supérieure à un seuil (B), on insère un silence supplémentaire.

6. Procédé selon l'une des revendications 4 et 5, dans lequel on remplace le silence supplémentaire par un bruit de fond (29).

7. Procédé selon la revendication 6, dans lequel le bruit de fond est à tonalité vocale.

8. Procédé selon l'une des revendications 1 à 7, dans lequel on élimine (20), de la restitution, les signaux mémorisés les plus anciens lorsque leur retard pour la restitution dépasse un seuil (C).

Patentansprüche

1. Verfahren zur Tonwiedergabe von Sprachsignalen (11, S1, 12, 13, S2), die in aufeinanderfolgenden Paketen (P1, P2, P3) empfangen werden, die aufeinanderfolgende Sprachzeitabschnitte darstellen und ursprünglich Pausen umfassen und vorübergehend gespeichert werden (4), bevor sie in Form von Tönen (5, 6) wiedergegeben werden, wobei das Verfahren **dadurch gekennzeichnet ist, daß** das Vorhandensein der Pausen (S1, S2) in den empfangenen Abschnitten erfasst und deren Wiedergabedauer derart reguliert wird, daß die Sprachsignale (11, 12), im Unterschied zu den Pausen, in einem Stück wiedergegeben werden.
2. Verfahren nach Anspruch 1, bei dem die Wiedergabe jeder Pause (S1), von der mindestens der Beginn der folgenden Tonsequenz verfügbar ist, abgekürzt wird (24, 28).
3. Verfahren nach Anspruch 1, bei dem die Wiedergabe der Pause (S1) abgekürzt wird (23, 28), wenn die Verzögerung bei der Wiedergabe der Sprachsignale eine Schwelle (A) überschreitet.
4. Verfahren nach einem der Ansprüche 1 bis 3, bei dem eine zusätzliche Pause (25, 27) eingefügt wird, wenn keine Folgesequenz (11) im Speicher verfügbar ist.

5. Verfahren nach einem der Ansprüche 1 bis 4, bei dem, wenn auf eine Pause (S1) im Speicher keine Tonsequenz mit einer größeren Dauer als eine Schwelle (B) folgt, eine zusätzliche Pause eingesetzt wird. 5
6. Verfahren nach einem der Ansprüche 4 und 5, bei dem die zusätzliche Pause durch ein Hintergrundgeräusch (29) ersetzt wird. 10
7. Verfahren nach Anspruch 6, bei dem das Hintergrundgeräusch eine vokale Tonbeschaffenheit aufweist
8. Verfahren nach einem der Ansprüche 1 bis 7, bei dem aus der Wiedergabe die ältesten gespeicherten Signale beseitigt werden (20), wenn ihre Verzögerung bei der Wiedergabe eine Schwelle (C) überschreitet. 15
20
7. Method according to Claim 6, whereby the background noise is at voice tone .
8. Method according to one of the Claims 1 to 7, whereby the oldest signals stored are eliminated (20), from the restoration, when their delay for restoration goes beyond a threshold (C).

Claims

1. Method of sound restoration of speech signals (11, S1, 12, 13, S2) received in successive packets (P1, P2, P3) representing successive temporal speech bursts containing originally periods of silence and stored temporarily (4) before being restored as sounds (5, 6), a process **characterised in that** the presence of the said silences (S1, S2) is detected, in the bursts received, and the duration of their restoration is regulated, in order to restore, all together, the speech signals (11, 12) other than the silences. 25
30
2. Method according to Claim 1, whereby there is a shortening (24, 28) of the restoration of any silence (S1) wherein at least the start of the next sound sequence is available. 35
3. Method according to Claim 1, whereby there is a shortening (23, 28) of the restoration of the silence (S1) if the delay in restoration of the speech signals goes beyond a threshold (A). 40
4. Method according to one of the Claims 1 to 3, whereby an extra silence (25, 27) is inserted when no following sequence (11) is available in the memory. 45
5. Method according to one of the Claims 1 to 4, whereby, when a silence (S1) is not followed in the memory by a sound sequence of a duration longer than a threshold (B), an extra silence is inserted. 50
6. Method according to one of the Claims 4 and 5, whereby the extra silence is replaced by a background noise (29). 55

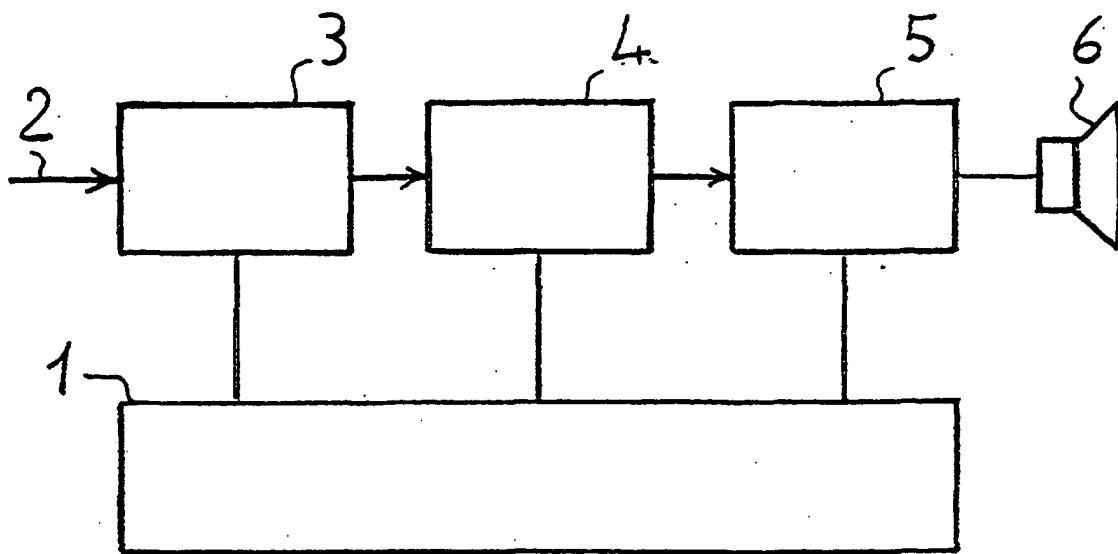
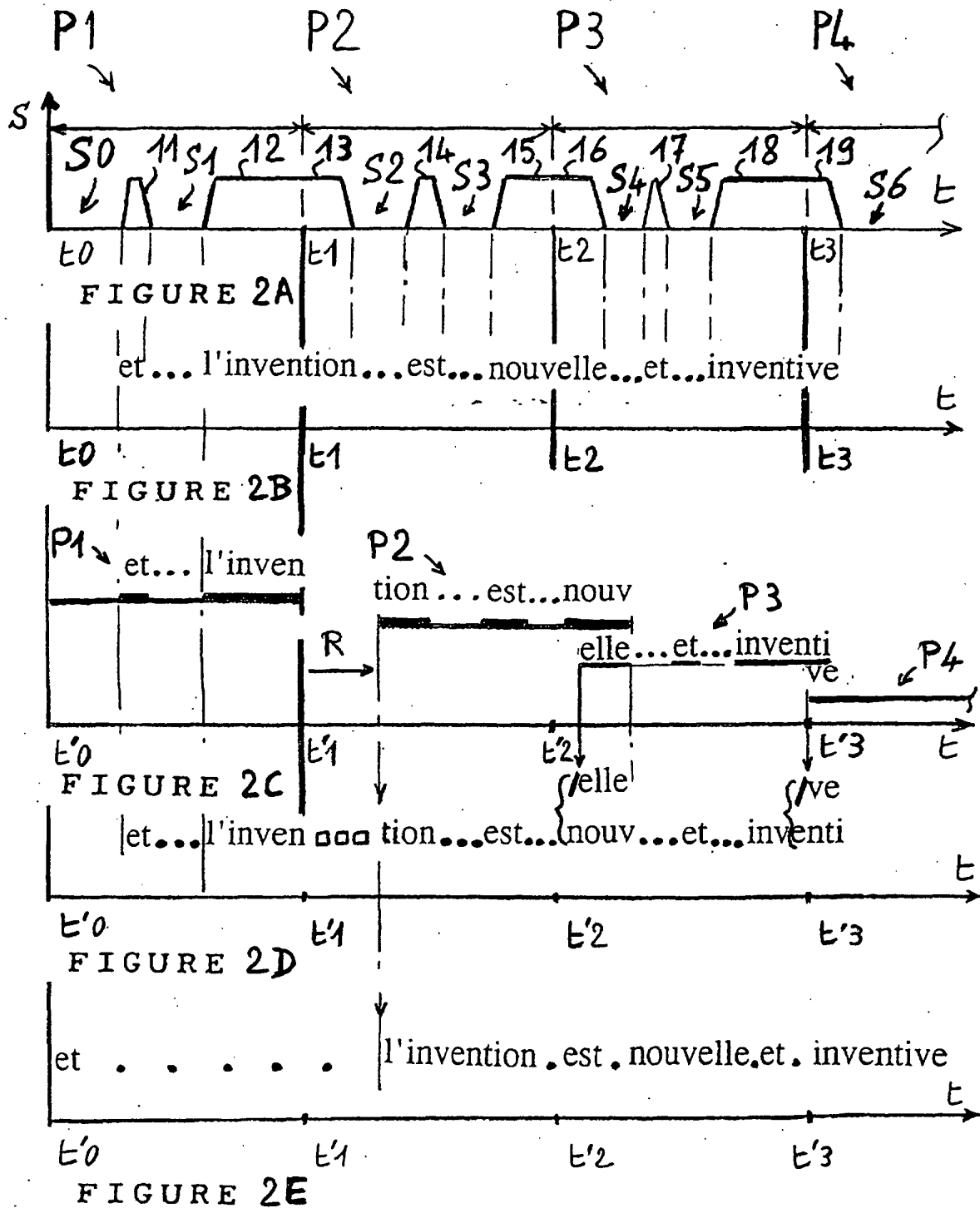


FIGURE 1



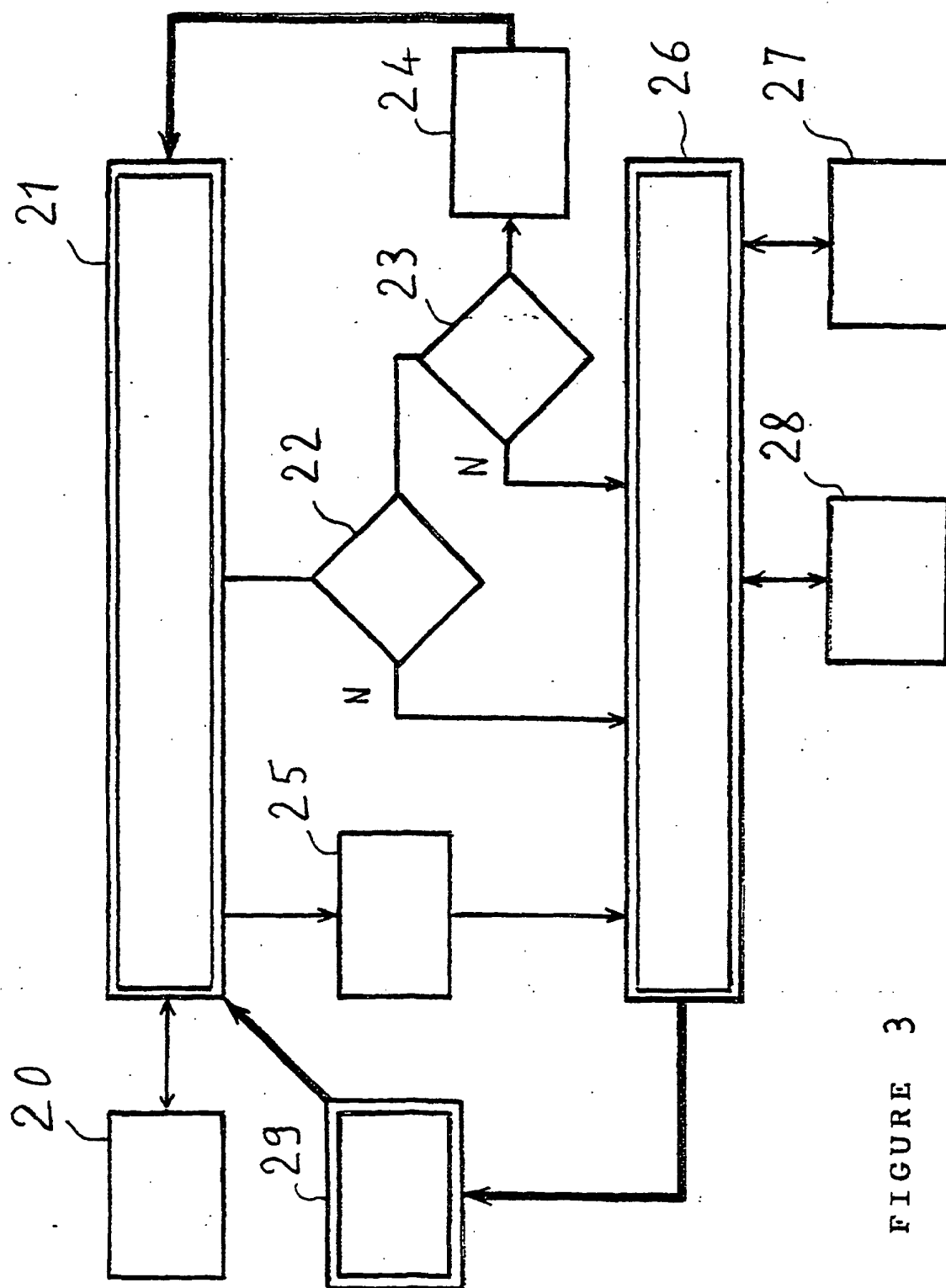


FIGURE 3