

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2148505

Российским агентством по патентам и товарным знакам на основании Патентного закона Российской Федерации, введенного в действие 14 октября 1992 года, выдан настоящий патент на изобретение

СПОСОБ ИДЕНТИФИКАЦИИ ФАКТА РЕЧЕВОЙ АКТИВНОСТИ ОПЕРАТОРА

Патентообладатель(и):

Закрытое акционерное общество "НЕЙРОКОМ"

по заявке № 99103468, дата поступления: 17.02.1999

Приоритет от 17.02.1999

Автор(ы) изобретения:

см. на обороте

Патент действует на всей территории Российской Федерации в течение 20 лет с 17 февраля 1999 г. при условии своевременной уплаты пошлины за поддержание патента в силе

Зарегистрирован в Государственном реестре изобретений Российской Федерации

г. Москва, 10 мая 2000 г.

Генеральный директор

А.Д. Корсакин



(51) МПК⁷ B60K28/06, G10L15/00

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ
СОБСТВЕННОСТИ,
ПАТЕНТАМ И ТОВАРНЫМ
ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

Статус: по данным на 27.11.2015 - действует

(21), (22) Заявка: **99103468/28, 17.02.1999**(24) Дата начала отсчета срока действия патента:
17.02.1999(45) Опубликовано: **10.05.2000**(56) Список документов, цитированных в отчете о
поиске: **US 5473726 A, 05.12.1995. RU 2111134
C1, 20.05.1998. US 5539859 A, 23.07.1996. US
5586215 A, 17.12.1996. EP 0660303 A1, 28.06.1995.**Адрес для переписки:
**101000, Москва, Старосадский пер. 8, НИЦ
"ЭЛДИС" РАН, Земляницыну М.А.**(54) **СПОСОБ ИДЕНТИФИКАЦИИ ФАКТА РЕЧЕВОЙ АКТИВНОСТИ ОПЕРАТОРА**

(57) Реферат:

Изобретение относится к диагностике психофизиологического состояния человека по речевому сигналу в процессе его профессиональной деятельности и может быть использовано при реализации контроля состояния водителя транспортного средства. Согласно данному способу регистрируют звуковое давление посредством двух разнесенных микрофонов, среднюю скорость воздушного потока - непосредственно перед ртом говорящего. Параметры движения видимой части рта определяют как мгновенное значение эффективной площади его раскрытия, а факт речевой активности устанавливают по критерию, использующему комбинацию упомянутых параметров, при этом пороговые значения для критерия подбирают экспериментально. Данный способ обеспечивает повышение надежности идентификации факта речевой активности оператора в условиях сильного зашумления и снижение требований к точности позиционирования датчиков дополнительной физиологической информации, а также микрофонов.

2 з.п. ф-лы, 2 ил.

(71) Заявитель(и):

**Закрытое акционерное
общество "НЕЙРОКОМ"**

(72) Автор(ы):

**Бабин Д.Н.,
Уранцев А.В.,
Мазуренко И.Л.,
Холоденко А.Б.**

(73) Патентообладатель(и):

**Закрытое акционерное
общество "НЕЙРОКОМ"**

Изобретение относится к способам диагностики психофизиологического состояния человека по речевому сигналу в процессе профессиональной деятельности и может быть использовано, в частности, в системах автоматического контроля состояния операторов различных профессий, преимущественно водителей транспортных средств.

Известно, что контроль функционального состояния оператора по речевому сигналу может быть эффективно использован для поддержания работоспособности, например, в условиях монотонии, когда оператор теряет бдительность (см. патент, RU 2111134 С1, НЕЙРОКОМ, В 60 К 28/06, 1998). Способ предусматривает ведение речевого диалога с последующим анализом адекватности реакции водителя (например, ответа в речевой форме) на смысловое содержание вопроса, для чего используется система распознавания речи, которая может быть реализована различными известными методами.

Известно также, что качество распознавания предъявляемого речевого сигнала определяется методами и средствами его приема и фиксации, приобретающими особое значение в реальных условиях работы оператора, особенно водителей транспортных средств, причем в этом случае стараются снизить общий уровень фона. Для этой цели известно использование нескольких микрофонов с совместной обработкой зарегистрированных каждым из них сигналов (патент, US 5539859, Robbe et al., G 10 L 9/00, 395/2.42, 1996).

При этом для эффективного распознавания и очистки от шума необходимо выявить только тот акустический сигнал, который и представляет собой речь данного оператора. Для такой цели известно использование дополнительных каналов физиологической информации, например, фиксирующих одновременно с речевым сигналом движение опорных точек на лице оператора с определением набора нормализованных векторов расстояний, описывающих мимику говорящего, с последующим их учетом (патент, US 5586215, Stork et al., G 10 L 5/06, 395/2.41, 1996) или других физиологических параметров, характеризующих эмоциональное состояние данного оператора (патент, EP 0660303 A1, AT&T Corp., G 10 L 9/16, 1995). Вместе с тем, в указанных источниках не содержится информации о возможности распознавания речи в условиях особо высокой зашумленности, например в кабине локомотива. Наиболее близким аналогом патентуемого способа является способ идентификации факта речевой активности оператора, включающий одновременную регистрацию звукового давления посредством установленного у рта говорящего первого микрофона, и параметров движения видимой части рта, сопоставление зарегистрированных параметров с пороговыми значениями и последующее выделение начала и конца речевого фрагмента, подлежащего распознаванию (патент, US 5473726, Marshall, G 10 L 5/06, 395/2.4, 1995). Способ может быть реализован посредством устройства, включающего микрофон и фотодатчик, подключенные к ЭВМ.

Однако вышеупомянутое изобретение не позволяет достаточно надежно идентифицировать факт речевой активности оператора в условиях сильного зашумления и вибрации, свойственным условиям работы ответственных профессий, например экипажей локомотивных бригад на железнодорожном транспорте. Технический результат изобретения состоит в повышении надежности идентификации факта речевой активности оператора в условиях сильного зашумления и снижении требований к точности позиционирования датчиков дополнительной физиологической информации, а также микрофонов.

Поставленная цель достигается тем, что способ идентификации факта речевой

активности оператора включает одновременную регистрацию звукового давления $P_1(t)$ посредством установленного у рта говорящего первого микрофона, и параметров перемещения элементов видимой части рта, сопоставление зарегистрированных параметров с пороговыми значениями и последующее выделение начала и конца речевого фрагмента, подлежащего распознаванию. Дополнительно регистрируют среднюю скорость $V(t)$ воздушного потока непосредственно перед ртом говорящего, а также среднее значение звукового давления $P_2(t)$ посредством второго микрофона, удаленного от первого. При этом в качестве параметров движения видимой части рта определяют мгновенное значение эффективной площади $S_{эфф}(t)$ его раскрытия, а факт речевой активности устанавливают при одновременном выполнении следующих условий:

$$|P_1(t) - P_2(t)| > E_1^{пор};$$

$$|P_1(t)| - |P_2(t)| > E_2^{пор};$$

$$|S_{эфф}(t)| > E_3^{пор};$$

$$|S'_{эфф}(t)| > E_4^{пор};$$

$$|V(t)| > E_5^{пор}(t),$$

где: $P_1(t)$, $P_2(t)$ - средние значения звукового давления, а $E_1^{пор}$, $E_2^{пор}$, $E_3^{пор}$, $E_4^{пор}$, $E_5^{пор}$ - пороговые значения, установленные экспериментально.

Кроме того, способ может характеризоваться тем, что средние значения звукового давления $P_1(t)$, $P_2(t)$ определяют путем усреднения соответствующих величин мгновенного звукового давления во временном интервале длительностью 10 - 100 миллисекунд.

Кроме того, способ может характеризоваться также тем, что среднюю скорость $V(t)$ воздушного потока непосредственно перед ртом говорящего определяют путем усреднения величины мгновенной скорости воздушного потока во временном интервале длительностью 10-100 миллисекунд.

В основе изобретения лежат следующие предпосылки, соображения и экспериментальные факты.

При наличии искомой речевой активности оператора среднее значение звукового давления $P_1(t)$ от ближнего микрофона должно превышать среднее значение звукового давления $P_2(t)$ от удаленного микрофона. Тем самым знак разности модулей этих сигналов $E_2(t) = |P_1(t)| - |P_2(t)|$ указывает направление распространения сигнала по оси, соединяющей ближний и дальний микрофоны, а абсолютная величина этой разности показывает относительную громкость источника сигнала в шуме. Кроме того, при отсутствии речевой активности оператора в предположении, что длина волны много больше расстояния между микрофонами, фазы шумовой составляющей сигнала незначительно изменяются при переходе от удаленного микрофона к ближнему, поэтому энергия $E_1(t) = |P_1(t) - P_2(t)|$ должна быть мала.

При анализе оптическим методом параметров движения видимой части рта сигналы, поступившие от фотоприемников, нормируют таким образом, что плотно закрытый рот соответствует нулевому уровню сигнала. Далее вычисляют модуль $E_3(t)$ сигнала с максимальным модулем производной и модуль производной этого сигнала $E_4(t)$. При

речевом ответе сигнал $E_3(t)$ должен быть отличен от константы в течение времени всего ответа; кроме того, модуль его производной определяет скорость движения губ. Использование этих параметров позволяет устройству не реагировать на "паразитные" движения губ оператора, например, при дыхании через рот. Регистрируемая информация о скорости $V(t)$ воздушного потока непосредственно перед ртом говорящего позволяет ввести дополнительные критерии для дифференциации входных данных, поскольку профили функции $V(t)$ во время дыхания и речи различны. Полученные на выходе значения сравнивают с пороговыми величинами, установленными экспериментально. Изобретение позволяет зафиксировать факт речевой активности, если в один и тот же момент времени все пять анализируемых величин превысили свои пороговые значения.

Сущность патентуемого способа поясняется чертежами, где:

на фиг. 1 изображена блок-схема устройства для реализации способа,

на фиг. 2 - алгоритм обработки сигналов для выделения факта речевой активности оператора.

Устройство содержит микрофон 1, располагаемый в непосредственной близости от рта оператора, микрофон 2, расположенный на некотором удалении (0,5-15 см) от первого микрофона, фотодатчик 3, предназначенный для регистрации параметров $S_{эфф}(t)$ движения видимой части рта и располагаемый перед губами оператора, а также датчик 4 средней скорости $V(t)$ воздушного потока непосредственно перед ртом говорящего, расположенный рядом с фотодатчиком 3.

Фотодатчик 3 включает инфракрасный излучатель 5 и приемное устройство 6, содержащее линейку фотоприемников. Оба микрофона и датчики монтируют на гарнитуре, одеваемой на голову оператора (на фигуре не показана). Введение нескольких фотоприемников позволяет снизить требования к точности позиционирования фотодатчика и увеличить свободу перемещения гарнитуры с датчиками без перенастройки устройства, что в свою очередь обеспечивает удобство при работе с устройством. Микрофоны 1,2, фотодатчик 3 и датчик 4 через индивидуальные аналого-цифровые преобразователи (АЦП) 7 - 10 подключены к блоку 11 анализа, который может быть реализован на базе компьютера.

Алгоритм обработки сигналов для выделения факта речевой активности оператора приведен на фиг. 2.

По команде "старт" (поз. 100) фиксируются сигналы $P_1(t)$ (поз. 102) и $P_2(t)$ (поз. 104) с первого и второго микрофонов, несущие информацию о звуковом давлении, а также сигнал о параметрах $S_{эфф}(t)$ движения видимой части рта (поз. 106) и средней скорости $V(t)$ воздушного потока непосредственно перед ртом говорящего (поз. 108). Далее вычисляется модуль разности сигналов звукового давления $E_1(t) = |P_1(t) - P_2(t)|$ (поз. 110). Далее, на следующем этапе (поз. 112) проводится проверка превышения величиной $E_1(t)$ порогового значения $E_1^{пор}$, установленного на этапе обучения системы. Если получен результат $E_1(t) < E_1^{пор}$, то фиксируется отрицательный результат. В противном случае на следующем шаге (поз. 114) вычисляется величина $Y_2(t)$, равная разности модулей указанных звуковых давлений. Далее проводится сравнение величин $E_2(t)$ и $E_2^{пор}$ (поз. 116), если $E_2(t) < E_2^{пор}$, то алгоритм дает отрицательный ответ, свидетельствующий о том, что идентифицируемая речь отсутствует. Если условие $E_2(t) > E_2^{пор}$ удовлетворяется, то алгоритм переходит к следующему шагу.

На следующем шаге (поз. 118) вычисляется величина $E_3(t) = |S_{эфф}(t)|$, а далее (поз. 120) проводится сопоставление этой величины с пороговым значением, установленным на этапе обучения системы. Если $E_3(t) > E_3^{пор}$, то производится вычисление величины $E_4(t)$, равной модулю производной величины $S_{эфф}(t)$ (поз. 122). На следующем этапе (поз. 124) производится проверка соответствия полученного значения $E_4(t)$ пороговой величине $E_4^{пор}$. Если удовлетворяется условие, то $E_4(t) > E_4^{пор}$, то проводится проверка выполнения условий для последнего анализируемого сигнала $E_5(t)$: модуля средней скорости $V(t)$ воздушного потока непосредственно перед ртом говорящего (поз. 126). В том случае, если не удовлетворяется условие $E_5(t) > E_5^{пор}(t)$ (поз. 128), делается общий вывод о том, что факт речевой активности отсутствует (поз. 130). Алгоритм при этом переходит к началу (поз. 100). Если же $E_5(t) > E_5^{пор}(t)$, то делается общий вывод о том, что анализируемый звуковой фрагмент представляет собой речь данного оператора. При этом началом звукового фрагмента является момент превышения всеми анализируемыми величинами соответствующих пороговых значений. Далее цикл анализа повторяется с дискретностью, определяемой частотой дискретизации блока анализа 11. Временной интервал при вычислении средних значений звукового давления $P_1(t)$, $P_2(t)$ и $V(t)$ воздушного потока непосредственно перед ртом говорящего составляет 10 - 100 миллисекунд.

Предлагаемое устройство позволяет повысить надежность идентификации факта речевой активности оператора в условиях сильного зашумления, а использование разнородных по физической природе датчиков - снизить требования к точности позиционирования гарнитуры на голове оператора.

Формула изобретения

1. Способ идентификации факта речевой активности оператора, включающий одновременную регистрацию звукового давления $P_1(t)$ посредством установленного у рта говорящего первого микрофона и параметров перемещения элементов видимой части рта, сопоставление зарегистрированных параметров с пороговыми значениями и последующее выделение начала и конца речевого фрагмента, подлежащего распознаванию, отличающийся тем, что дополнительно регистрируют среднюю скорость $V(t)$ воздушного потока непосредственно перед ртом говорящего, а также среднее значение звукового давления $P_2(t)$ посредством второго микрофона, удаленного от первого, при этом в качестве параметров движения видимой части рта определяют мгновенное значение эффективной площади $S_{эфф}(t)$ его раскрытия, а факт речевой активности устанавливают при одновременном выполнении следующих условий:

$$|P_1(t) - P_2(t)| > E_1^{пор},$$

$$|P_1(t)| - |P_2(t)| > E_2^{пор},$$

$$|S_{эфф}(t)| > E_3^{пор},$$

$$|S'_{эфф}(t)| > E_4^{пор},$$

$$|V(t)| > E_5^{пор}(t),$$

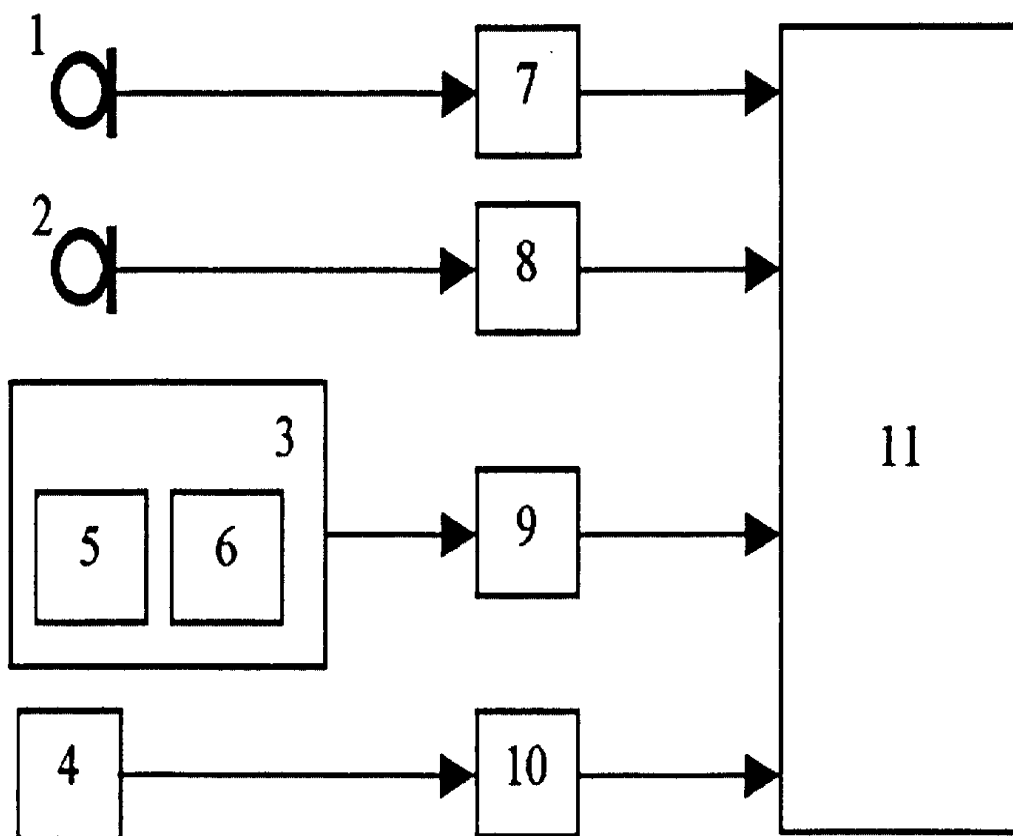
где $P_1(t)$, $P_2(t)$ - средние значения звукового давления;

$E_1^{пор}$, $E_2^{пор}$, $E_3^{пор}$, $E_4^{пор}$, $E_5^{пор}$ - пороговые значения, установленные экспериментально.

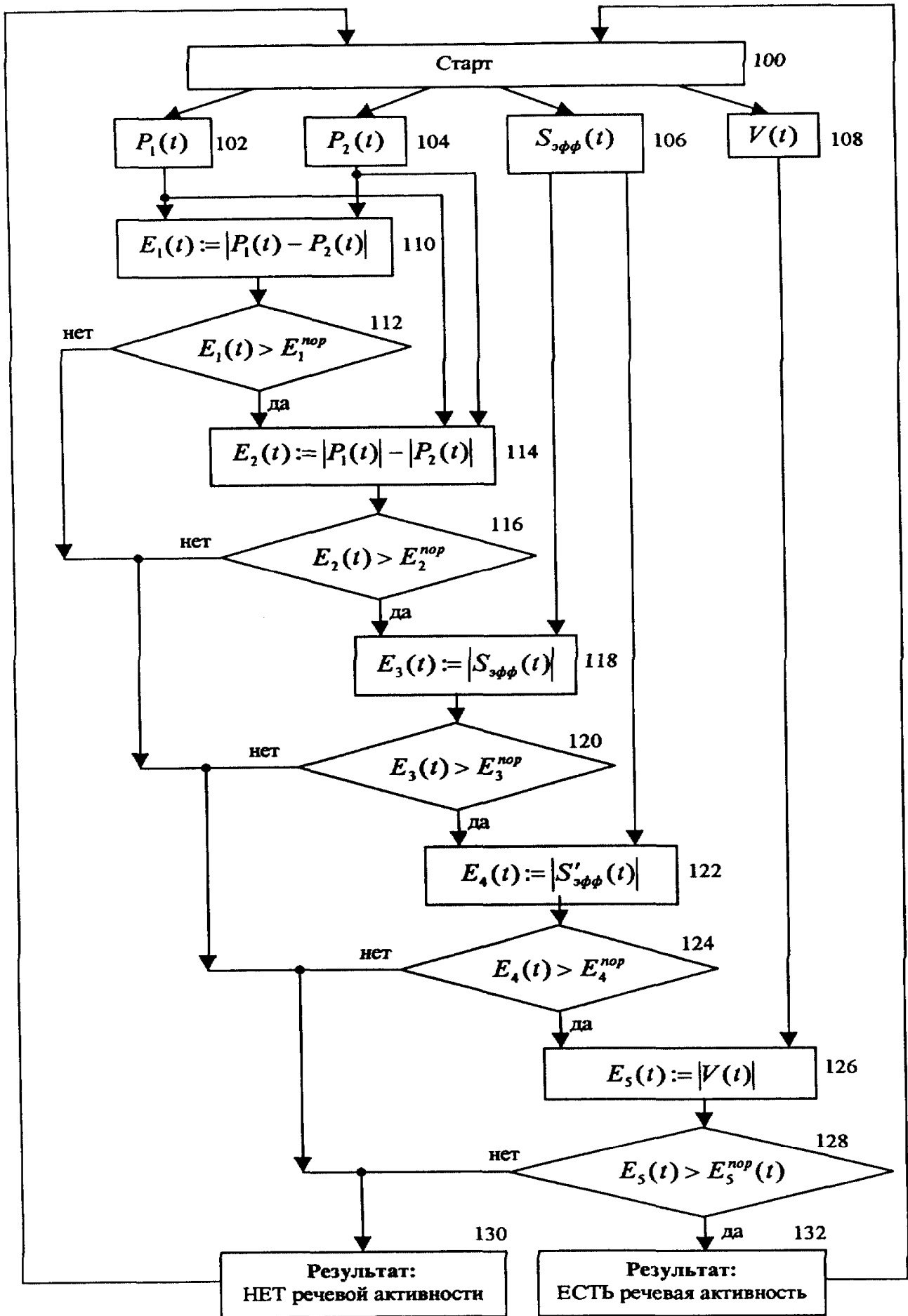
2. Способ по п. 1, отличающийся тем, что средние значения звукового давления $P_1(t)$, $P_2(t)$ определяют путем усреднения соответствующих величин мгновенного звукового давления во временном интервале длительностью 10-100 мс.

3. Способ по п. 1 или 2, отличающийся тем, что среднюю скорость $V(t)$ воздушного потока непосредственно перед ртом говорящего определяют путем усреднения величины мгновенной скорости воздушного потока во временном интервале длительностью 10-100 мс.

РИСУНКИ



Фиг. 1



Фиг.2