Claims

1. A hearing aid system based on the fusion of vision and hearing, characterized by comprising an environment monitoring module, a multidimensional recognition module, and an intelligent generation module;

the environmental monitoring module consists of a noise monitoring unit, an electromagnetic monitoring unit, and an environmental analysis unit, the noise monitoring unit is used to obtain noise data of the current environment and number the obtained current environmental noise data to form a noise dataset, the electromagnetic monitoring unit is used to obtain electromagnetic data of the current environment and number the obtained current environmental electromagnetic data to form an electromagnetic dataset, the noise monitoring unit and electromagnetic monitoring unit send the numbered noise dataset and electromagnetic dataset to the environmental analysis unit through the network, the environmental analysis unit calculates the environmental impact index based on the received dataset and determines whether the current environment will affect the speech recognition effect according to the environmental impact index, if it is determined that the current environment will affect the speech recognition effect, it sends an environmental anomaly signal to the multi-dimensional recognition module;

the multidimensional recognition module includes a speech recognition unit and a lip recognition unit, the speech recognition unit is used to generate speech recognition text based on speech recognition, the lip recognition unit is used to generate fused recognition text based on lip recognition combined with speech recognition after receiving abnormal signals, the speech recognition unit and lip recognition unit send the generated recognition text to the intelligent generation module;

the intelligent generation module is used to display the final text.

2. The hearing aid system based on the fusion of vision and hearing according to claim 1,

characterized in that the noise monitoring unit is connected to a microphone through a network and obtains noise data through the microphone.

3. The hearing aid system based on the fusion of vision and hearing according to claim 2, characterized in that the electromagnetic monitoring unit is connected to an electromagnetic environment monitoring instrument through a network and obtains electromagnetic data through the electromagnetic environment monitoring instrument.

4. The hearing aid system based on the fusion of vision and hearing according to claim 3, characterized in that the expression for the number of the noise dataset is: Zs_1、Zs_2、Zs_3、…、Zs_n, where Zs_1 represents the first noise data obtained through the microphone, Zs_n represents the last noise data obtained through the microphone, and n represents the total number of data in the dataset.

5. The hearing aid system based on the fusion of vision and hearing according to claim 4, characterized in that the expression for the number of the electromagnetic dataset is: Cs_1、Cs_2、Cs_3、…、Cs_n, where Cs_1 represents the first electromagnetic data obtained through an electromagnetic environment monitoring instrument, Cs_n represents the last electromagnetic data obtained through an electromagnetic environment monitoring instrument, n represents the total number of data in the dataset, and the data acquisition time of the noise dataset is consistent with that of the electromagnetic dataset.

6. The hearing aid system based on the fusion of vision and hearing according to claim 5, characterized in that the calculation formula for the environmental impact index $HJyx$ is:

$$HJyx = \alpha_1 * \frac{Zs_i}{Zs_o} + \alpha_2 * \frac{Cs_i}{Cs_o}$$

in the calculation formula, $HJyx$ represents the environmental impact index, Zs_i represents the i-th noise data, Cs_i represents the i-th electromagnetic data, Zs_orepresents the maximum noise interference data allowed by the speech recognition system, Cs_o represents the maximum electromagnetic interference data allowed by the speech recognition system, α_1 represents the weight of noise factors, and α_2 represents the weight of electromagnetic factors.

7. The hearing aid system based on the fusion of vision and hearing according to claim 6, characterized in that when the calculated value of the environmental impact index is greater than the environmental impact threshold, it represents that the current environment will affect the speech recognition effect, and sends an abnormal signal to the multidimensional recognition module.

8. The hearing aid system based on the fusion of vision and hearing according to claim 7, characterized in that the speech recognition text generation method is:

S1.1 the speech recognition unit processes the speech signal into frames based on the speech recognition system, and extracts D-dimensional acoustic features from each frame to obtain a feature sequence. The expression of the acoustic feature sequence is:

$$X = [X_1、X_2、X_3、\cdots、X_t]$$

in the expression, X represents the acoustic feature sequence, X_1 represents the D-dimensional feature vector of frame 1, X_t represents the D-dimensional feature vector of frame t, and t represents the number of time frames.

S1.2 the speech recognition unit calculates the conditional probability of the acoustic feature sequence on the candidate phoneme sequence through an acoustic model, and selects the text sequence with the highest probability as the generation result, the calculation formula for the probability of the acoustic feature is:

$$S_1 = \prod_{i=1}^{t} S(X_i|Q_i)$$

in the calculation formula, S_1 represents the probability of acoustic features, X_i represents the D-dimensional feature vector of the i-th frame, and Q_i represents the phoneme state corresponding to the i-th frame; $\prod_{i=1}^{t} S(X_i|Q_i)$ represents the multiplication of all probabilities from frame 1 to frame t.

9. The hearing aid system based on the fusion of vision and hearing according to claim 8, characterized in that the method for generating fused recognition text is:

S2.1 the lip recognition unit uses face detection technology to locate the lip area, extracts the D-dimensional visual feature vector of each lip image frame, and obtains a visual feature sequence, the expression of the visual feature sequence is:

$$V = [V_1、V_2、V_3、\cdots、V_t]$$

in the expression, V represents the visual feature sequence, V_1 represents the D-dimensional visual feature vector of frame 1, $V_t$represents the D-dimensional visual feature vector of frame t, and t represents the number of time frames.

S2.2 the lip recognition unit calculates the conditional probability of visual features on candidate phoneme sequences through a visual acoustic model, and the probability calculation formula for the visual features is:

$$S_2 = \prod_{i=1}^{t} S(V_i|Q_i)$$

in the calculation formula, S_2 represents the probability of visual features, V_i represents the D-dimensional feature vector of the frame, and Q_i represents the phoneme state corresponding to the frame; $\prod_{i=1}^{t} S(V_i|Q_i)$represents the multiplication of all probabilities from frame 1 to frame t.

S2.3, the lip language recognition unit calculates the fusion probability between the probability of acoustic features and the probability of visual features, and selects the text sequence with the highest probability as the generated result, the calculation formula for the fusion probability is:;

$$S_r = \beta_1 \cdot S_1 + \beta_2 \cdot S_2$$

in the calculation formula, S_r represents the fusion probability, β_1 represents the weight of the probability of acoustic features, β_2 represents the weight of the probability of visual features, $\beta_1 + \beta_2 = 1$.

10. The hearing aid system based on the fusion of vision and hearing according to claim 9, characterized in that the intelligent generation module is internally connected to a display device, and the final text is displayed through the display device.