

THIS OPINION WAS NOT WRITTEN FOR PUBLICATION

The opinion in support of the decision being entered today (1) was not written for publication in a law journal and (2) is not binding precedent of the Board.

Paper No. 18

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

Ex parte DAVID G. STORK,
GREGORY J. WOLFF and EARL I. LEVINE

Appeal No. 94-3601
Application 07/889,619¹

ON BRIEF

MAILED

APR 9 1996

PAT. & T.M. OFFICE
BOARD OF PATENT APPEALS
AND INTERFERENCES

Before THOMAS, KRASS and BARRETT, Administrative Patent Judges.
KRASS, Administrative Patent Judge.

DECISION ON APPEAL

This is a decision on appeal from the final rejection of claims 1 through 8 and 10 through 17, constituting all the claims remaining in the application.

¹ Application for patent filed May 26, 1992.

Appeal No. 94-3601
Application 07/889,619

The invention is directed to a speech recognition system wherein facial features (i.e., lip reading) are used in conjunction with acoustic data in a neural network classification system.

Representative independent claim 1 is reproduced as follows:

1. A speech recognition system for recognizing utterances belonging to a pre-established set of allowable candidate utterances using acoustic speech signals and selected concomitant dynamic visual facial feature motion between selected facial features associated with acoustic speech generation, comprising:

a. an acoustic feature extraction apparatus for converting signals representative of acoustic speech into a corresponding acoustic feature vector set of signals;

b. a dynamic visual feature extraction apparatus for converting signals representative of the selected concomitant dynamic visual facial feature motion associated with acoustic speech generation into a corresponding visual feature vector set of signals; and

c. a time delay neural network classifying apparatus for generating a conditional probability distribution of the allowable candidate speech utterances by accepting and operating on a set of current and time delayed dynamic acoustic feature and visual feature vector sets respectively supplied by the acoustic and visual feature extraction apparatus.

The examiner relies on the following references:²

Baji et al. (Baji) 5,163,111 Nov. 10, 1992

² Although the examiner refers to the reference Petajan (U.S. Patent No. 4,975,960) in the grounds of rejection and in various places throughout the answer, it is clear from the entire record and the fact that no new ground of rejection is entered in the answer that the examiner meant to refer to Pentland.

Appeal No. 94-3601
Application 07/889,619

Waibel et al. (Waibel), "Phoneme Recognition: Neural Networks vs. Hidden Markov Models," IEEE, (September 1988), pp.107-110.

Pentland et al. (Pentland), "Lip Reading: Automatic Visual Recognition of Spoken Words," M.I.T. Media Lab Vision Science Technical Report 117, (January 1989), pp.1-9.

Claims 1, 6 and 11 through 17 stand rejected under 35 U.S.C. 102(e) as anticipated by Baji. Claims 2 through 5, 7, 8 and 10 stand rejected under 35 U.S.C. 103. As evidence of obviousness, the examiner cites Baji with Waibel with regard to claims 2 through 5 and Baji with Pentland with regard to claims 7, 8 and 10.³

Rather than reiterate the arguments of appellants and the examiner, reference is made to the briefs and answer for the respective details thereof.

OPINION

We will not sustain the rejection of claims 1, 6 and 11 through 17 under 35 U.S.C. 102(e) or the rejection of claims 2 through 5, 7, 8 and 10 under 35 U.S.C. 103.

With regard to claim 1, we agree with the examiner that Baji discloses an acoustic feature extraction apparatus and a dynamic visual feature extraction apparatus, as broadly claimed. Clearly, elements 31, 32 and 34 may constitute the claimed acoustic feature extraction apparatus and image interface unit 21

³ The examiner includes claim 9 in this rejection but claim 9 has been cancelled and should form no part of the rejection.

Appeal No. 94-3601
Application 07/889,619

may constitute the claimed dynamic visual feature extraction apparatus. Appellants' argument that circuit 5 of Baji is not an acoustic feature extraction apparatus and that element 21 is not a visual feature extraction unit is unpersuasive. First, circuit 5 need not form part of the acoustic feature extraction apparatus. In any event, appellants' argument appears to be one of semantics. In other words, no matter what one calls these elements of Baji, they seem to provide the same signals as broadly claimed.

Further, we would even be willing to go so far as to agree with the examiner that there is some type of time delay in Baji, either inherent, or at least suggested. However, we still will not sustain the rejection of claim 1 under 35 U.S.C. 102(e) because claim 1 calls for more than merely a time delay neural network (TDNN) classifying apparatus having a time delay. The claim is very specific that the TDNN must generate a "conditional probability distribution" of allowable speech utterances and this generation is based on an operation upon "a set of current and time delayed dynamic acoustic feature and visual feature vector sets."

Appellants refer to this in their arguments at the bottom of page 5 of the principal brief wherein they state, referring to Baji,

Appeal No. 94-3601
Application 07/889,619

No enabling method is described as to how the necessary delay is to be determined and applied, such as Applicants [sic, Applicants'] time delay neural network of claim 1, item (c) [emphasis in the original].

However, the examiner never comes to grips with this claimed limitation and we find no hint or suggestion in Baji as to operation by a TDNN upon a set of current and time delayed dynamic acoustic feature and visual feature vector sets in order to generate a conditional probability distribution of allowable candidate speech utterances, as claimed.

With regard to independent claim 16, while this claim does not recite the acoustic feature extraction apparatus, it does require a TDNN for generating a conditional probability distribution of allowable candidate utterances from the set of visual feature vector signals generated over a prescribed interval of time. The examiner has not come to grips with this claimed limitation and therefore, for the reasons supra, with regard to the rejection of claim 1 under 35 U.S.C. 102(e), we also will not sustain the rejection of claim 16 under 35 U.S.C. 102(e).

We have reviewed the Waibel and Pentland references but since they provide no suggestion or reason for modifying the teaching of Baji to provide for at least the deficiencies noted supra with regard to claims 1 and 16, we also will not sustain

Appeal No. 94-3601
Application 07/889,619

the rejection of claims 2 through 5, 7, 8 and 10 under 35 U.S.C.
103.

The examiner's decision is reversed.

REVERSED



JAMES D. THOMAS
Administrative Patent Judge)



ERROL A. KRASS
Administrative Patent Judge)

BOARD OF PATENT
APPEALS AND
INTERFERENCES



LEE E. BARRETT
Administrative Patent Judge)

Appeal No. 94-3601
Application 07/889,619

Blakely, Sokoloff, Taylor and Zafman
12400 Wilshire Boulevard
7th Floor
Los Angeles, CA 90025-1026