CLAIMS

1. A computer-implemented method comprising:

receiving audio streams captured by a plurality of distributed devices;

comparing the audio streams captured by the plurality of devices to determine whether the audio streams are representative of sound from a same meeting;

based on the comparing indicating the audio streams are representative of sound from the same meeting, establishing a meeting instance for an intelligent meeting;

identifying a user corresponding to a distributed device of the plurality of distributed devices in the intelligent meeting;

determining a preferred language of the user;

generating, by a hardware processor, a transcript from the received audio streams as the intelligent meeting is occurring;

translating the transcript into the preferred language of the user to form a translated transcript as the intelligent meeting is occurring; and

providing the translated transcript to the distributed device as the intelligent meeting is occurring.

2. The method of claim 1, wherein providing the translated transcript comprises providing the transcript with translated text for display on the distributed device as the intelligent meeting is occurring.

3. The method of claim 1, wherein providing the translated transcript comprises converting text of the translated transcript to speech for output to the user of the distributed device as the intelligent meeting is occurring.

4. The method of claim 1, wherein providing the translated transcript comprises providing speaker identities for each translated utterance of the transcript.

5. The method of claim 1, wherein the determining the preferred language of the user comprises accessing a user preference previously established for the user indicating the preferred language.

6. The method of claim 1, wherein the comparing the audio streams captured by the plurality of devices to determine that the audio streams are representative of sound from the same meeting comprises:

calculating normalized cross correlation coefficients between signals of the audio streams; and

determining whether a predetermined threshold is transgressed, wherein the predetermined threshold being transgressed indicates that the audio streams are representative of sound from the same meeting.

7. The method of claim 1, further comprising:

performing continuous speech separation on the received audio streams from the plurality of distributed devices to separate speech from different speakers speaking at the same time into separate audio channels, the generating the transcript being based on the separated audio channels.

8. The method of claim 1, wherein identifying the user comprises:

receiving a video signal capturing the user; and

matching a stored image of the user with the video signal to identify the user.

9. The method of claim 1, wherein identifying the user comprises:

matching a stored voice signature of the user with speech from the audio streams.

10. The method of claim 1, wherein identifying the user comprises:

obtaining a user identifier associated with the distributed device.

11. A non-transitory machine-storage medium having instructions for execution by a processor of a machine to cause the processor to perform operations comprising:

receiving audio streams captured by a plurality of distributed devices;

comparing the audio streams captured by the plurality of devices to determine whether the audio streams are representative of sound from a same meeting;

based on the comparing indicating the audio streams are representative of sound from the same meeting, establishing a meeting instance for an intelligent meeting;

identifying a user corresponding to a distributed device of the plurality of distributed devices in the intelligent meeting;

determining a preferred language of the user;

generating a transcript from the received audio streams as the intelligent meeting is occurring;

translating the transcript into the preferred language of the user to form a translated transcript as the intelligent meeting is occurring; and

providing the translated transcript to the distributed device as the intelligent meeting is occurring.

12. The machine-storage medium of claim 11, wherein providing the translated transcript comprises providing the transcript with translated text for display on the distributed device as the intelligent meeting is occurring.

13. The machine-storage medium of claim 11 wherein providing the translated transcript comprises converting text of the translated transcript to speech for output to the user of the distributed device as the intelligent meeting is occurring.

14. The machine-storage medium of claim 11, wherein providing the translated transcript comprises providing speaker identities for each translated utterance of the transcript.

15. The machine-storage medium of claim 11, wherein the determining the preferred language of the user comprises accessing a user preference previously established for the user indicating the preferred language.

16. The machine-storage medium of claim 11, wherein comparing the audio streams captured by the plurality of devices to determine that the audio streams are representative of sound from the same meeting comprises:

calculating normalized cross correlation coefficients between signals of the audio streams; and

determining whether a predetermined threshold is transgressed, wherein the predetermined threshold being transgressed indicates that the audio streams are representative of sound from the same meeting.

17. The machine-storage medium of claim 11, wherein the operations further comprise:

performing continuous speech separation on the received audio streams from the plurality of distributed devices to separate speech from different speakers speaking at the same time into separate audio channels, the generating the transcript being based on the separated audio channels.

18. The machine-storage medium of claim 11, wherein identifying the user comprises:

receiving a video signal capturing the user; and

matching a stored image of the user with the video signal to identify the user.

19. The machine-storage medium of claim 11, wherein identifying the user comprises:

matching a stored voice signature of the user with speech from the audio streams.

20. A device comprising:

one or more hardware processors; and

a memory device coupled to the processor and having a program stored thereon that, when executed by the one or more hardware processors, causes the one or more hardware processors to perform operations comprising:

receiving audio streams captured by a plurality of distributed devices;

comparing the audio streams captured by the plurality of devices to determine whether the audio streams are representative of sound from a same meeting;

based on the comparing indicating the audio streams are representative of sound from the same meeting, establishing a meeting instance for an intelligent meeting;

identifying a user corresponding to a distributed device of the plurality of distributed devices in the intelligent meeting;

determining a preferred language of the user;

generating a transcript from the received audio streams as the intelligent meeting is occurring;

translating the transcript into the preferred language of the user to form a translated transcript as the intelligent meeting is occurring; and

providing the translated transcript to the distributed device as the intelligent meeting is occurring.