June 2020

CAMERA-BASED POSITIONAL AUDIO METADATA GENERATION FOR ON-STAGE PRESENTER

Øivind Stuan

Christian Theien

CAMERA-BASED POSITIONAL AUDIO METADATA GENERATION FOR ON-STAGE PRESENTER

AUTHORS:
Øivind Stuan
Christian Theien

ABSTRACT

Techniques are described herein to enable a voice to dynamically track the physical position of an on-stage presenter. This may be accomplished without using additional sensors or complex configurations.

DETAILED DESCRIPTION

Techniques are described herein for enabling and simplifying the setup and use of automatic mixing of audio from an on-stage presenter to match the physical location of the presenter. Thus, dynamically positioned audio may be supported based on the movement of an on-stage presenter.

A known spatial relationship between the loudspeaker system and some visual anchors visible to the camera view may be leveraged. These anchors could be part of the video signal being shown if the screen is within the field of view of the camera. In the case of an integrated system, the spatial relationship between the screen and loudspeakers is known, however for custom installations this would need to be configured. This custom configuration could be as simple as arranging the location of the loudspeaker system relative to the video screen (or some other visual tracking anchor) in a web-based graphical configuration tool.

If there is only one presenter to be tracked, the only required configuration may involve indicating to which microphone signal(s) to apply the tracking. If there are multiple simultaneous presenters to be tracked, this could be handled using face recognition to aid full-body detection. The microphone input channel assignment may be performed by speech recognition (e.g., "Hi, my name is John Doe," "and my name is Jane Doe."). A manual fallback is to assign the individual user names to each microphone input.

Figure 1 below illustrates an example system in which there are multiple presenters.

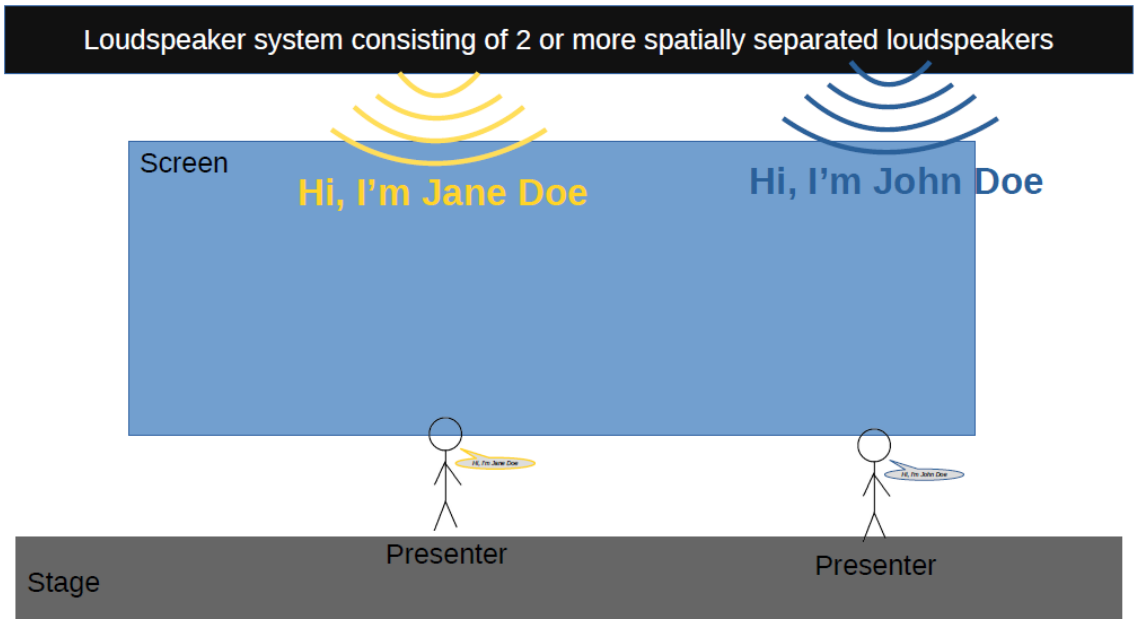

Figure 1

Figure 2 below illustrates an example system in which the location of the presenter is tracked.

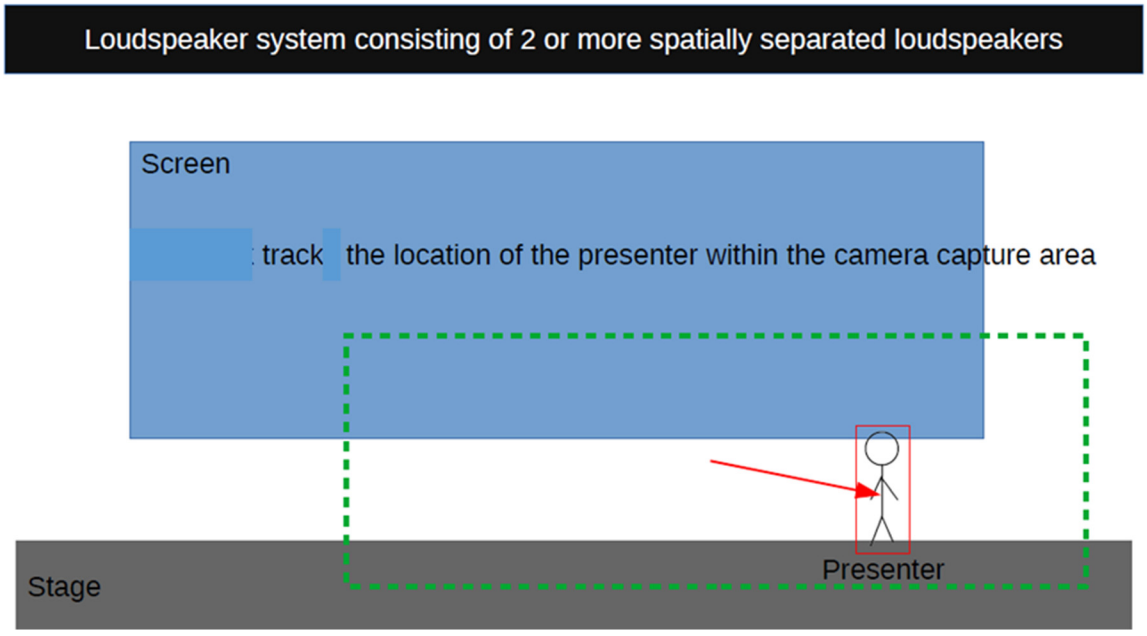

Figure 2

Figure 3 below illustrates an example system having a known or configured spatial relationship between the main screen and the loudspeakers.

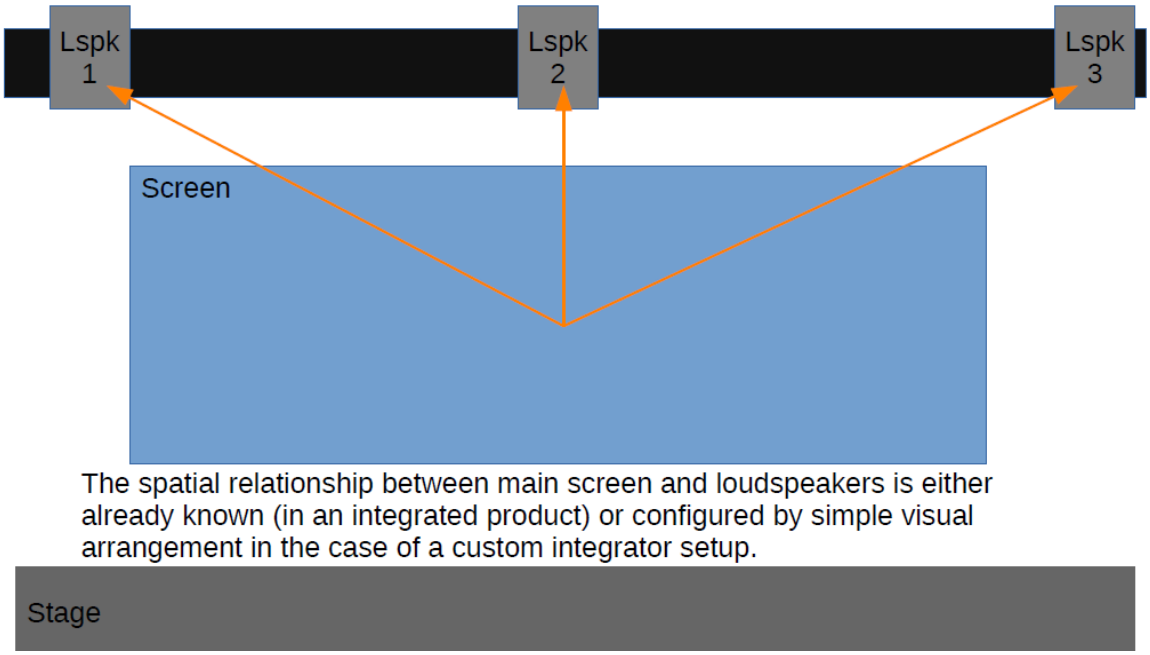

Figure 3

Figure 4 below illustrates an example system having an inferred spatial relationship between the camera capture area and the screen.

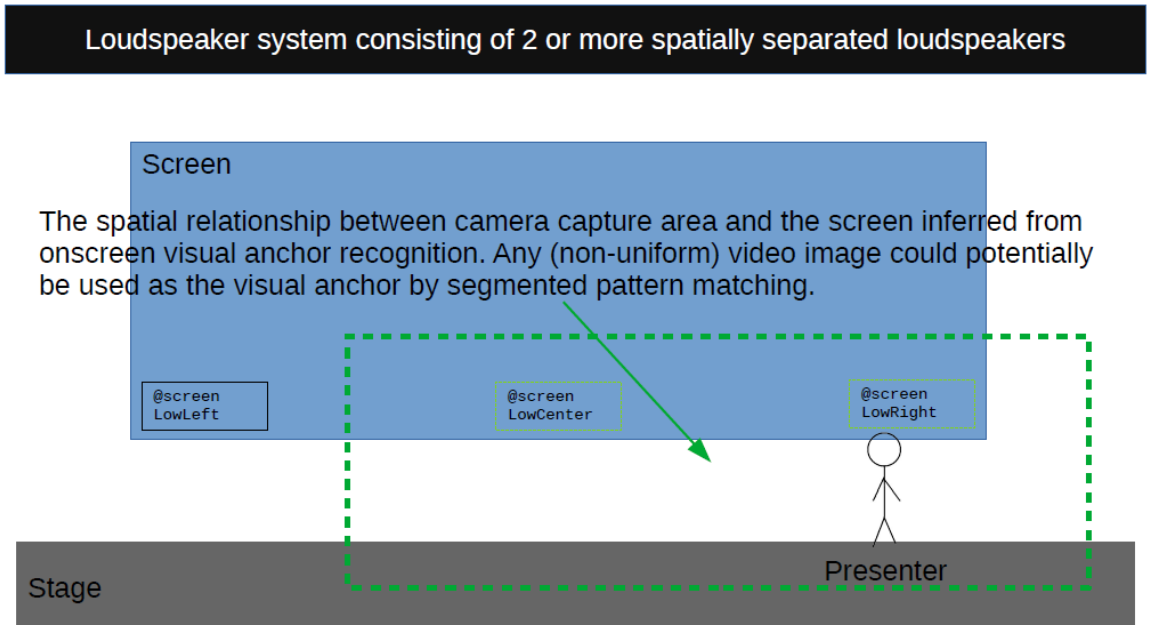

Figure 4

Figure 5 below illustrates an example system in which the location of the presenter is determined based on one or more spatial relationships.

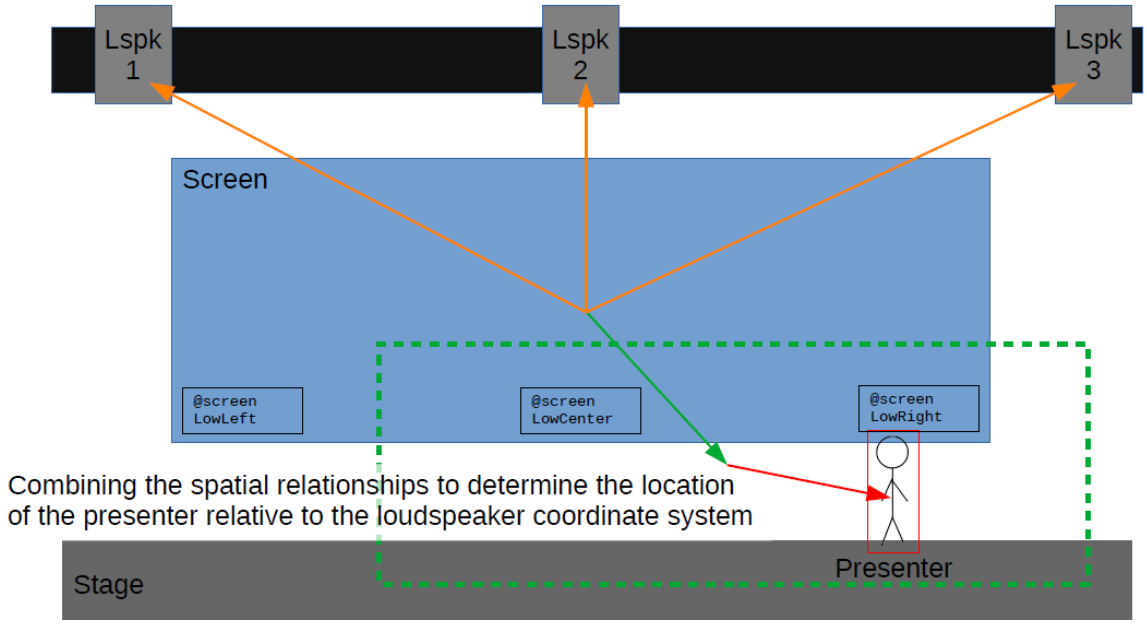

Figure 5

Figure 6 below illustrates an example system in which microphone input channels are assigned based on speech recognition.

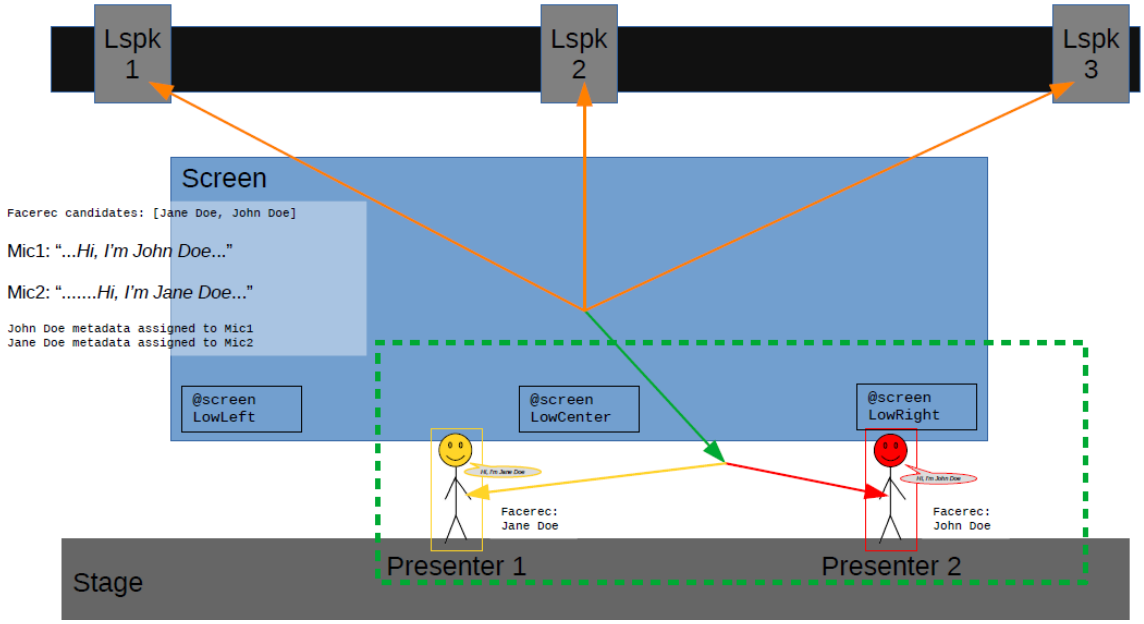

Figure 6

In summary, techniques are described herein to enable a voice to dynamically track the physical position of an on-stage presenter. This may be accomplished without using additional sensors or complex configurations.