Several factors will need to be counterbalanced. To keep an overview I have studied the paper and present the following factors:

For a single set of stimuli (e.g. 'dofoe' and 'pima'):
    If the familirisation set should start with 'bal' or 'boek'.
    If it is infant directed or adult-directed speech.
    Which visual and audio stimuli are linked (e.g. 'dofoe' to image1 and 'pima' to image2).
    On which side a picture should be shown (i.e. left or right).

Each factor above has two levels. Together with two set of audio stimuli this makes for 32 possible configurations. Since each configuration consists of 5 phases this would entail changing 160 definition of trials. To lower this number the configuration is divided between participant variables and the participant group.

Participant Fields:
'condition': IDS or ADS
Determines if the infant directed or adult-directed speech audio stimuli should be used.
'first_famil_picture': ball or book
Determines if the familiarization should start with the book or with the ball.

Different groups:
Three factors are set with the group selection:
1. Which set of stimuli words should be used (i.e. [modi and dofoe] or [doboe and pima]).
2. Which image should be associated with the first word of the set (i.e. image1 or image2).
3. Which side this image should be displayed.

The group name is determined as follows:
    <SIDE><IMAGE><SET> (e.g. LEFT1MODI)
Where *SIDE* is either LEFT or RIGHT
Where *IMAGE* is either 1 or 2
Where *SET* is either DOBOE or MODI.
There are *__eight__* groups.