

Experiment 1.

In this Experiment, we selected turns from spontaneous German conversations (Lindenstrasse Corpus; Kohler, 1996). The sound file labelled NAT contains one of these turns. We manipulated the original turn in three ways. The sound file labelled NPITCH was resynthesized using Praat (Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer [Computer program]. Version 5.3.73*. Retrieved April 21, 2014, from <http://www.praat.org/>), so that the pitch was set to the average pitch in the original recording (PSOLA resynthesis). The sound file labelled NWORDS was low-passed filtered at 500Hz, also using Praat. Finally, the sound file labelled MARY, was produced by a speech synthesizer (MARY, Schröder, M., & Trouvain, J. (2003). The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *Int. Journal of Speech Technology*, 6, 365-377.).

Experiment 2.

In this Experiment, the sound file labelled MARY was superimposed on videos of a virtual agent performing different types of gaze movements. The virtual agent used is EMBR (Heloir, A., & Kipp, M. (2009). EMBR - A realtime animation engine for interactive embodied agents. *In Proc. of the 9th Int'l Conference on Intelligent Virtual Agents*. Berlin Heidelberg: Springer). In the video labelled NOGAZE, the virtual agent maintains mutual gaze with the viewer throughout the turn (there are no changes in gaze direction). In the video labelled NATURAL, the agent looks away at the start of the video and then looks back to the viewer about 600 ms before the end of the spoken turn. In the video labelled AVERTED, the timing of the gaze movements is the same as in the NATURAL video, expect that the agent does not looks back towards the viewer at the end, but shifts her gaze to the opposite direction while maintaining averted gaze. In the video labelled, EARLY, finally, the sequence of movements is the same as in the video labelled NATURAL, expect that the final movement back towards the viewer occurs around 1600 ms before the end of the spoken utterance.