1. Introduction

By following others people's eye gaze, we can recognize their focus of attention. The ability to coordinate attention to an object or event with others is a phenomenon called joint attention (Mundy et al., 1986). Although joint attention is thought to be built upon basic neural mechanisms detecting the gaze direction, it also extends to understanding pointing or other social cues by the interactive partner or sharing an awareness of the object or event with the partner; therefore, it is considered to be a uniquely human ability underlying social communication. It is widely known that people attend to each other (by eye contact, smiling, mimicking facial expressions and gestures etc.) to augment shared understanding, and it plays a crucial role in verbal communication. Joint attention is important for at least two reasons: 1) it allows us to learn about the other person's inner state (such as interest, emotion and intentions), 2) it informs us about what the speaker is talking about and where the object of interest is in the environment.

Autism spectrum disorders (ASD), including autistic and Asperger's disorders, are characterized by qualitative impairments in social interaction (American Psychiatric Association [APA], 2000). One of the important features of these social impairments is thought to be a deficit of joint attention (APA, 2000; Mundy et al., 1986). A lack of joint attention has been well documented in the clinical literature and recently it has drawn attention as an early marker of ASD (APA, 2000); however, contrary to the clinical findings, experimental studies generally have reported intact joint attention in ASD (Chawarska, Klin, & Volkmar, 2003; Kylliäinen & Hietanen, 2004; Rutherford & Krysko, 2008; Senju et al., 2004; Swettenham et al., 2003; Vlamings et al., 2005; see Nation et al., 2008 for review). These studies applied Posner's cueing paradigm (Posner, 1980), in which uni-modal (i.e. visual) cue-target pairs were used. In the paradigm, subjects first saw a gaze cue (directed toward right or left), followed by a target, a dot or a letter, which appeared either on the right or left of the display screen. The subjects were asked to locate the target and respond as quickly and accurately as possible, and the reaction