Originally, I had prepared a study by Ciccone et al. (1979) for you about parallel speech and gestures restrictions by aphasics, but I decided to introduce you a more progressive research (considering used methods/methodology) in So, Kita and Goldin-Meadow's article. As I have noticed, this study examines healthy people, not aphasics, and they basically claim that to explore substitution of speech by gestures is relevant particularly by people in “unstable cognitive state“ = also by aphasic.

So, Kita and Goldin-Meadow solve a question whether speakers use gesture to help them specify referents when they fail to do so in speech. Thus, they investigate how healthy speakers semantically coordinate speech and gesture to disambiguate information that is crucial for discourse processing.

They proceed from Kita and Özyürek (2003), where is established so-called interface hypothesis. That means, when speech fails to uniquely specify a referent, gesture should also fail to identify the referent.

As you have noticed, they found out that gesture was rarely used to compensate for the absence of lexical specificity in pronouns or nouns. BUT! You must take into consideration that only nine participants were involved in this research, so it would be quite “dangerous“ to make generalization for all healthy people based on this particular research.

According to them, speakers did use gesture to specify the identity of a referent. However, speakers did not use gesture to specify the identity of a referent unless that referent was also uniquely specified in speech – which means speakers did not use gesture to compensate for the under-specification found in their speech, but they used gesture in parallel with their speech.

So, Kita and Goldin-Meadow also mention few examples from different languages to support their theory, e.g. that English has a word to convey an arced trajectory (“swing”) but Turkish and Japanese do not and they even do not use a gesture to compensate this lexical gap – they use speech when referring to changing of location and a gesture with straight trajectory – not with arced trajectory. Still, this quite interesting observation could not be considered for a generally valid argument as well.

Finally, I would like to thank you for some inspiring questions/idea like e.g. this research is not dealing with a more complicated situation ”when the speaker is describing multiple referents and their location in space and in relation to each other”. I think results would be not so much corresponding with above mentioned ”hand-in-hand” hypothesis.

Considering co-speech gestures, there is an agreement across literature that hands are temporally synchronized and semantically integrated with speech within this type of gestures. If you need to say something verbally, you cannot recall a concrete word so you only use pronouns like “this” or “that” together with a gesture describing a concrete shape or movement, so it is called “pro-speech” gesture (it helps you to recall a concrete word so you are able to say it verbally afterwards).