Peter Robinson

When your sat-nav says 'right', why do you turn left?

In November 1983, a Colombian jumbo jet en route from Paris to Bogotá was making a scheduled stop in Madrid. Landing in the dark, the crew made a mistake with the instrument landing system, turned on to an incorrect track and flew into a hill. An analysis of the cockpit voice recorder revealed that some minutes before the crash, an audible ground proximity warning system had told the crew, "Pull up! Pull up!" The pilot replied "Shut up, gringo." Those were his last words. All 20 crew and 161 of the 172 passengers were killed.

More recently, BMW in Germany had to recall all the cars equipped with its innovative navigation system. The system was excellent technically, providing highly accurate information about the car's location and routes to arbitrary street addresses. However, the system gave information and instructions using a female voice. German male drivers (the vast majority of BMW's customers) do not take directions from women. The result was described as a "gender fender bender".

Clearly there is more to voice output from computers than simply presenting clear and accurate information. Issues such as ethnicity, gender, personality and manner affect our response to speech, even when that speech is obviously synthesised by a machine. In *Wired for Speech*, Clifford Nass and Scott Brave explore the psychology of human voice perception and its implications for the design of computer interfaces using recorded speech and speech synthesis.

The book arises from a ten-week course that Nass ran at Stanford University to ex-

Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship

By Clifford Nass and Scott Brave

MIT Press
297pp, £20.95
ISBN 0 262 14092 6

plore the design and implementation of voice interfaces. The students (including Brave) were divided into groups who designed and conducted experiments to test various psychological theories about voice interfaces. *Wired for Speech* presents quantitative results derived from 20 of these experiments run with thousands of human subjects. Their story is scientifically compelling and humanly fascinating.

The experiments simulated various scenarios. For example, users' reactions to racial or regional accents were tested using an e-commerce website on which four product descriptions were read by an agent with either an Australian or Korean accent. The users were then asked questions about the products' likeability and the descriptions' credibility. Participants were also asked about the agent's overall quality. The results confirmed earlier studies indicating that the agents were rated much more positively when they spoke in an accent that matched the user's ethnicity. Participants also evaluated the products and descriptions more positively when they were spoken by an agent whose accent matched their own. There is a case for matching the accent of spoken warnings in aircraft to those of their pilots if the warnings are to be taken seriously.

Another effect also becomes important as computer power and network speed increase, allowing images to accompany speech. Photographs of the agents were also shown, but randomly chosen to be ethnically Australian or Korean. The participants' perceptions of the agents' racial appearance in the photograph did not affect their judgment in the same way as the accent. Social identification through accent is a much stronger influence than racial identification. However, the participants gave consistently lower ratings to agents whose voices and photographs were inconsistent. They also rated the products and their descriptions more highly when the voices and photographs were consistent. Social identification and stereotyping is important in human interaction, but it is weakened by inconsistency.

Social identification also carries through to regional accents. For example, speakers with British accents are perceived in the US to be more intelligent, which give British academics a useful advantage at conferences. Boston accents are perceived as elitist, and Southern US accents are perceived as more egalitarian. These labels are powerful influences, and some Japanese call centres are already using the area codes of callers to select voices with matching regional accents.

Nass and Brave write from a US perspective and mention class and status only in passing. It would be interesting to repeat their experiments in a British context. What do we really make of call centre operators with accents from Newcastle or Calcutta?

A similar analysis can be applied to gender in voices. Despite huge advances in speech synthesis in recent years, voices are still clearly artificial with curious accents, phrasings and mispronunciations. Nevertheless, one experiment showed that people clearly identified with a voice that "matched" their gender. Participants were presented with a dilemma requiring a choice, and were offered advice by a computer-generated voice that was randomly chosen to be "female" or "male". Female participants found the female voice to be more trustworthy, and male participants found the male voice to be more trustworthy.

The same experiment also confirmed many women's experience: when they say something, they are ignored, but when a man says the same thing, people pay attention. There is a simple effect from the gender of the voice. Men listening to a male voice were most convinced, and men listening to a female voice were least convinced. This innate prejudice extends even to synthetic voices. BMW may have used a female voice in its navigation system for excellent technical reasons to do with audibility of higher pitched voices, but it failed to take account of learned social behaviours and assumptions.

Nass advised BMW on the redesign of its voice interface, using the same rigour that it already applied to other aspects of car design. Apart from being male, what other characteristics should the voice assume? Should it use the "loud, dominant slightly unfriendly voice that suggested a German automotive engineer"? Should it be a dominant pilot or a subservient chauffeur? Should it be a friend or a back-seat driver? The final choice was the voice of a co-pilot, indicating a high level of competence and moderate assertion but still friendly. Although other experiments show that recorded voice should use the first person, the BMW adviser confirms its subordinate role by avoiding the use of "I".

However, there is still the question of the adviser's mood. Voice emotion influences driving performance. Another experiment considered simulated interactions between drivers who were happy or upset with voices that were enthusiastic or subdued. Drivers who interacted with voices that matched their own emotional state had less than half as many accidents on average as drivers who interacted with mismatched voices. Again, the important factor is alignment and consistency rather than absolute mood. Assessing the emotional state of a driver is not easy, but it is clearly important. Our research into this topic at the computer laboratory in Cambridge is making contributions that are already exciting interest from car manufacturers.

Wired for Speech is lucidly written, and it has copious references to the original source material for those who wish to assess the evidence or undertake more research. It should be compulsory reading for anyone involved in human-computer interaction, not just for those using synthesised or recorded speech.

However, the book will appeal to a much wider readership. A baby in the womb is aware of its own mother's voice and can distinguish her from other speakers. Our opinions of other people are moulded by listening to their voices for just a few moments. *Wired for Speech* explains these basic psychological phenomena and reveals patterns of human behaviour that are crucial to our lives as communicating beings.

Peter Robinson is professor of computer technology, Cambridge University.