Research Proposal **Mark Scott** **PIN 307872**

The research I propose is based on observations made during prior experimental work on the McGurk effect (McGurk & MacDonald: 1976 Nature 264). This effect causes people to 'mishear' speech sounds when the sound is spliced with a mismatched video. For example, when the sound /ba/ is dubbed over video of a face saying /ga/, subjects 'hear' /da/. The video causes people's categorization of the acoustics of /ba/ to be temporarily shifted to the category /da/. During previous research, I noticed that this categorization-shift sometimes persisted even after the video trigger was removed. In fact, through extended exposure to the McGurk effect my own categorization of some sounds became 'stuck' in the shifted condition. This persistence of the McGurk effect has never been demonstrated in a controlled experiment and to my knowledge has never been mentioned in the literature. I propose to investigate if and under what conditions the McGurk illusion persists after the visual trigger is removed.

This research is related to the question of *speaker variability*. One way to understand speaker variability is to consider the differences in pronunciation between adults and children. There are differences of degree - the lower pitch of the adult voice; and there are differences of kind – the resonating chambers in the adult vocal tract are not only larger, but of different proportion. Thus, the acoustics of an adult's voice are NOT simply a 'scaled up' version of a child's. There are also individual differences in pronunciation and physical morphology, independent of age, that cause acoustic variability between speakers. Yet, despite all this variability, listeners categorize different people's productions of a word like "hello" as being essentially the same.

Listeners may handle speaker-variability by using prior experience as a guide to categorization – this is one theory of speech-perception. The persistence of the McGurk effect would demonstrate exactly this: that current speech-perception is influenced by previous experience (in this case exposure to the McGurk effect). I will conduct experiments to show whether and under what conditions a permanent shift in speech-sound categorization can be induced by exposure to the McGurk effect.

In addressing how humans overcome speaker-variability, this research is valuable to the field of automated speech-recognition, which must deal with the same problem. This research will also provide experimental data to help decide between the two main theoretical approaches to speech perception.

Theoretical background

There are two competing theories of how speech perception deals with speaker variability: the *abstractionist* theory and the *episodic memory* theory (Goldinger: 1998 Psychological Review 105).

The *abstractionist* theory claims that mental representations of speech sounds are highly abstract and invariant. In this view of perception, invariant properties of a speech sound are extracted by filtering out the acoustic differences, which are considered noise. This process of eliminating the differences between a token of a sound and its abstract mental representation is known as "speaker normalization". Of course, the acoustic differences may serve other purposes such as determining a speaker's age, sex, emotional state and identity.

The *episodic memory* theory views speaker variability as a source of information, rather than noise. In this view, representations of speech sounds are not abstract sets of features with the noise filtered out, but collections of memory traces of the entire signal, 'noise' included. Speech sounds are identified by comparing them against the memory traces to determine the best fit.

A demonstration of the persistence of the McGurk effect would be strong evidence in support of the episodic memory theory. A null result would be consistent with the abstractionist theory.

Methodology

Subjects will be trained with large numbers of McGurk stimuli (with the face and voice of the stimuli kept constant so that they become familiar). According to the abstractionist view, training will have no effect on the perception of audio-only stimuli. However, if the episodic theory is correct, subjects will build memory traces of the speaker's voice/pronunciation and when the visual influence is removed they will continue to categorize sounds based on their training – categorizing what they heard as /ba/ before training as /da/ after training.

Mark Scott **PIN 307872**

1. Contributions to research

Articles published in refereed journals

a. **Scott, M**. (2006) Puns of mythical proportions: Catullus 95. Research Reports of Kobe-City Technical College 44: 89-92.

b. **Scott, M**. (2006) No support for the Motor Theory of speech perception from sensorimotor priming. Research Reports of Kobe-City Technical College 44: 93-96.

c. **Scott, M.**, K. Dohlus, and G. Pintér. (2005) 時間知覚における視聴覚情報の相互作用 (Seeing geminates and hearing singletons). Theoretical and Applied Linguistics at Kobe Shoin 8: 133-142.

Non-refereed contributions

d. **Scott, M**. (2005) Does ambiguity trigger all compatible sound representations? Proceedings of the 131st General Meeting of the Linguistic Society of Japan: 180-185.

e. **Scott, M.** and K. Dohlus. (2005) Visual duration-information is used in language perception. Proceedings of the 19th General Meeting of the Phonetic Society of Japan: 61-66.

f. **Scott, M.** (2000) n/y alternation in Mushuau Innu. Memorial University Aldrich Conference.

2. Most significant contributions to research - My most important contributions are those labelled b, d and e above. The first, ('b' above) tested the *Motor Theory*, which argues that in speech-perception we perceive the articulatory gestures of the speaker, recovered through the acoustic effects of these gestures; so, perception is achieved through the mechanisms of production. While it is true that the brain must have both perceptual and motor representations of sounds, and that these representations must be linked (as is shown when we imitate sounds, where we convert the sounds we hear into motor commands), I do not believe the Motor Theory as it stands is tenable. If perception and articulation are achieved through the same neural mechanisms, then, by extension, it may be possible for articulation to induce interference effects on perception. This experiment looked for such effects, but could find none; though some of the findings indicated that further research would be worthwhile.

Subjects were asked to hold their breath at their lips. This would engage the same muscles as those used to produce /p/, and would create sensory feedback consistent with /p/ (raised air-pressure in the mouth, lip compression, raised velum, lack of vibrotactile sensation in the lips). If perception relies on production, then a possible consequence could be that while a subject's articulators are in the configuration for /p/, perception of /p/ will be affected; an effect would be evidence for the Motor Theory. A comparison condition was used in which subjects held their breath at the glottis. This would be equally distracting, but would not create sensorimotor feedback consistent with /p/.

As expected, there was no significant difference in response times to /p/ when subjects held their breath at their lips from when they did so at their glottis. However some of the results were close enough to statistical significance to warrant further investigation.

Experiment 'd' above investigated the effect of ambiguity on phonological representations. An ambiguous sound is compatible with more than one percept; during perception, the less compatible candidates must be dropped in favour of the most compatible. It has been shown that in this process, the semantic representations of rejected candidates are triggered. For example, "time", with the voice-onset-time (VOT) of the initial /t/ trimmed so that it is slightly consistent with /d/ (though still perceived as /t/), will trigger the semantic network of "dime". Thus, "time" (with reduced VOT) will speed processing of words like "penny", which are semantically related to the rejected candidate "dime"; this is semantic priming. My experiment investigated the phonological level: are rejected candidates primed or inhibited at this level? If the discarding of reject candidates is achieved by suppression, there should be slower response-times (inhibition) at the phonological level.

I used the McGurk effect to create the reject-candidates. The McGurk effect is an illusion created when vision and hearing disagree over the identity of a sound. For example, audio of /bi/ spliced with video of a face pronouncing /gi/ is perceived as /di/. Thus, subjects hear a sound acoustically consistent with /bi/ but must reject this candidate. If this rejection is achieved by suppression, then perception of a

following audio-only /bi/ should be inhibited, and response times to this following /bi/ should be slower.

The results of this experiment were inconclusive because the audiovisual mismatch inherent to the McGurk effect slowed response times so much that any delay caused by inhibition would have been drowned out and undetectable. However, the question that prompted the experiment is an important one, so I am designing a new experiment that will eliminate the confounding factor of audiovisual mismatch.

I feel my most important contribution to date is 'e' in the list above. This experiment tested if visual duration-information is involved in speech perception. Most previous tests of the McGurk effect examined shifts in perception from one place-of-articulation to another (as in the example above, where the acoustics of /bi/ were perceived as /di/). Since place-of-articulation is signalled (visually) by changes in facial shape, these experiments were testing the influence of visual shape-perception on speech perception. In our experiment, we tested if visual duration-perception could also influence speech perception. To test this, we relied on the contrast between long and short /p/ in Japanese (a contrast which does not exist in English). There is little difference between these sounds in facial shape; the significant difference is the duration of the labial closure; thus visual duration-perception must be used.

It is important to note that shape and duration perception are separate forms of visual perception, processed differently and presumably processed by different areas of the brain (though the mechanisms of visual duration-perception are not well understood). The fact that visual shape-perception affects speech perception does not necessarily imply that visual duration-perception is used in the same way.

Our experiment confirmed that visual duration-perception is indeed used in speech-perception. Subjects auditory perception of long or short /p/ was influenced by the accompanying video.

I was the lead author, so I was responsible for the concept and design of the experiment as well as the writing and presentation of the article for the conference. My co-author and I divided up the remaining work: the literature review and the running of the experiment on subjects.

3. Applicant's statement - During my master's degree I received a thorough grounding in phonological theory and in conducting an extended research project. After graduation I worked for two years as an English teacher in Japan. I then received a two-year research-scholarship from the Japanese government. During the scholarship I was essentially left to pursue whatever area of research interested me without supervision. As a result, I was forced to learn a lot of new skills. I developed skills in computer programming (to create the software for my experiments), statistical analyses, speech synthesis, acoustic analysis software and video editing. I also learned how to handle the sorts of problems that are encountered when dealing with human subjects.

When my research scholarship ended, there were several research projects that I wanted to pursue, so I remained in Japan to complete them, paying my way by teaching at three universities.

I have a lot of experience in teaching and tutoring. This started in high school, where I worked as a math-tutor. In my first year of university, my dormitory organized tutorials for those falling behind and I was asked to be the tutor for history and philosophy classes. When I started in linguistics, I continued to volunteer as a tutor in a variety of subjects. During my MA I worked as a tutor in my capacity as a teaching-assistant; since then I have been continuously employed as a teacher, either part or full-time.

I have also done volunteer work aside from tutoring. Most recently I organized a charity concert in Japan to raise money for the victims of the Indonesian tsunami. As well as organizing the event, I performed in one of the groups (as the guitar/trumpet/mandolin/violin player of a Latin-music band). We raised more than 240 000 yen (c. $2600 CDN) for the International Red Cross.

Outside of my academic life, my main interests are music and theatre. I play a variety of instruments and I have been in bands and played for theatrical productions for many years. I have worked as an actor (amateur and professional) since high school. In fact, over the last academic year I worked as a drama instructor at Setsunan University. I believe my experience in acting complements my career in academics, improving my skills as a teacher and an oral presenter.