

The relations between articulatory precision and sensory acuity

Phil Hoole

The relations between articulatory precision and sensory acuity

Mainly a review, but with some ideas for future investigations, and with special reference to cochlear implant speech.

Sensory: auditory, somatosensory, visual

Linking auditory acuity and speech production (1)

The first MIT studies

Perkell, J., Guenther, F. , Lane, H., Matthies, M., Stockmann, E., Tiede, M., Zandipour, M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, 116, 2338-2344.

Contrast in production estimated from EMA and formant data.
EMA example: '**cod**' vs. '**cud**'

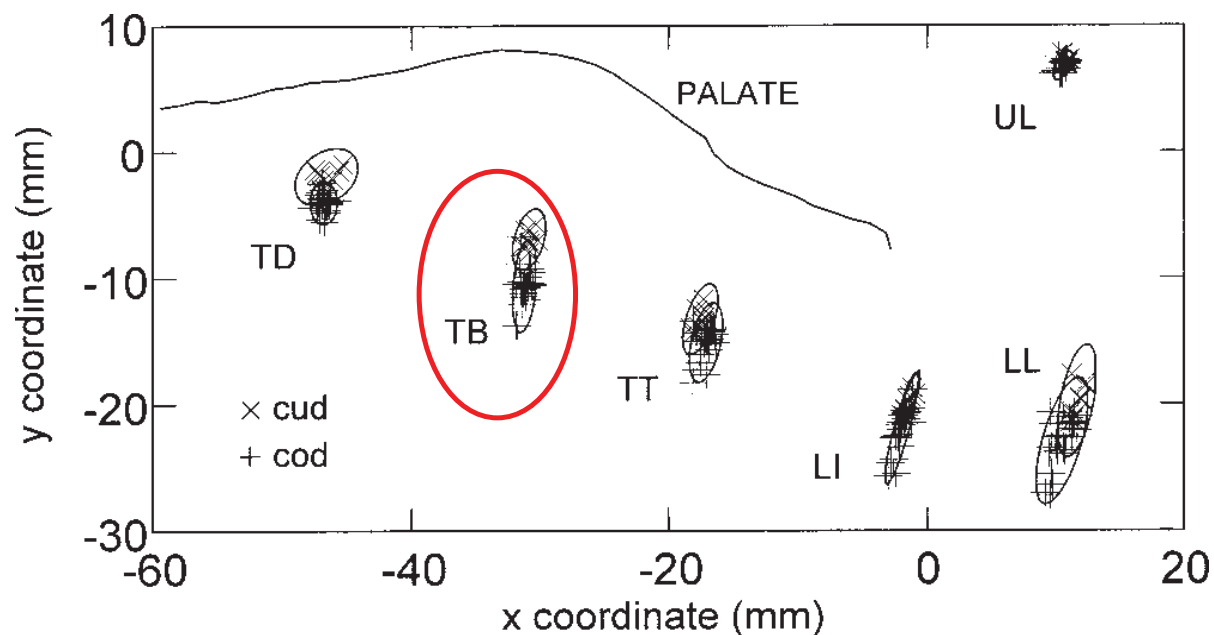
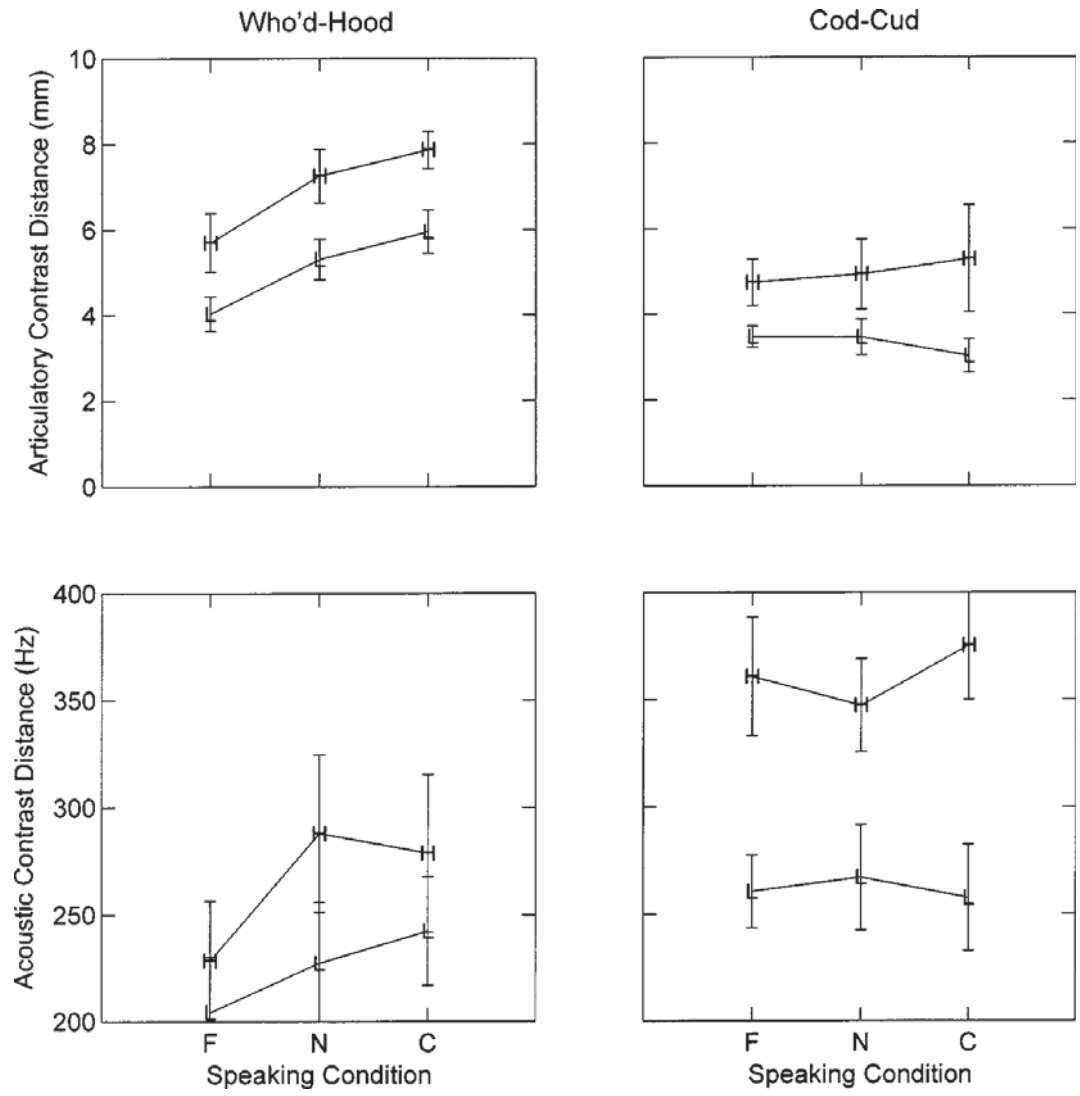


FIG. 1. EMMA transducer coil locations during all the productions of tokens containing the /ɑ/ in *cod* (+) and the /ʌ/ in *cud* (×) by one female subject in the normal condition. The transducer coils are located at points on the tongue dorsum (TD), tongue blade (TB), tongue tip (TT), lower incisor, (LI), lower lip (LL), and upper lip (UL). The midsagittal palatal contour (up to about the dento-alveolar junction on the right) is shown for reference.

Perceptual acuity estimated from ABX discrimination test
(7-step continua from 'cod' to 'cud' and 'who'd' to 'hood')
Note: continuum with only a small number of steps is not ideal.

Based on results, subjects divided into high discriminators and low discriminators



Contrast distance in production for High and Low discriminators in perception

FIG. 4. Articulatory contrast distance (for tongue body position—upper panel) and acoustic contrast distance (for separation in the formant plane—lower panel) as a function of the three speaking conditions. The left-hand panel gives results for *who'd-hood*, the right for *cod-cud*. Findings for high discriminators (labeled “H”) and low discriminators (“L”) are plotted separately. Error bars are one standard error about the mean.

Linking auditory acuity and speech production (2)

Basically the same story for fricatives.

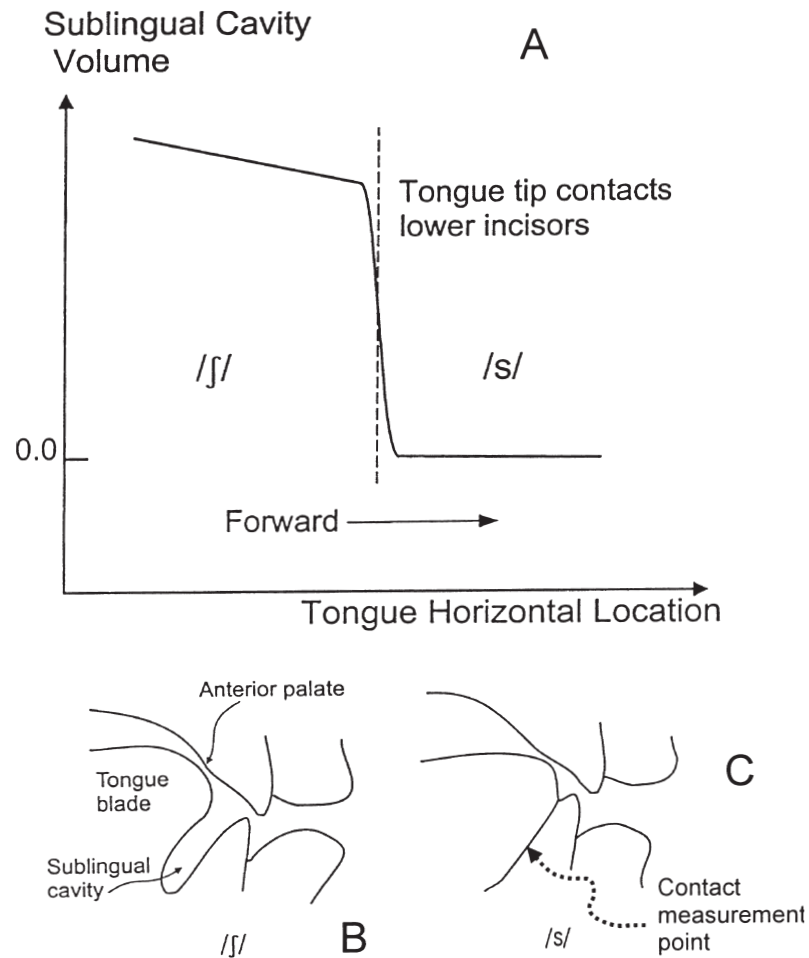
Contrast in production based on spectral centre of gravity for /s/ and /ʃ/.

Acuity in perception based on ABX discrimination test with **s-ʃ** continuum.

But also a first extension to somatosensory aspects

Perkell, J., Matthies, M., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., Guenther, F. (2004b). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of speech, language, and hearing research*, 47(6), 1259-69.

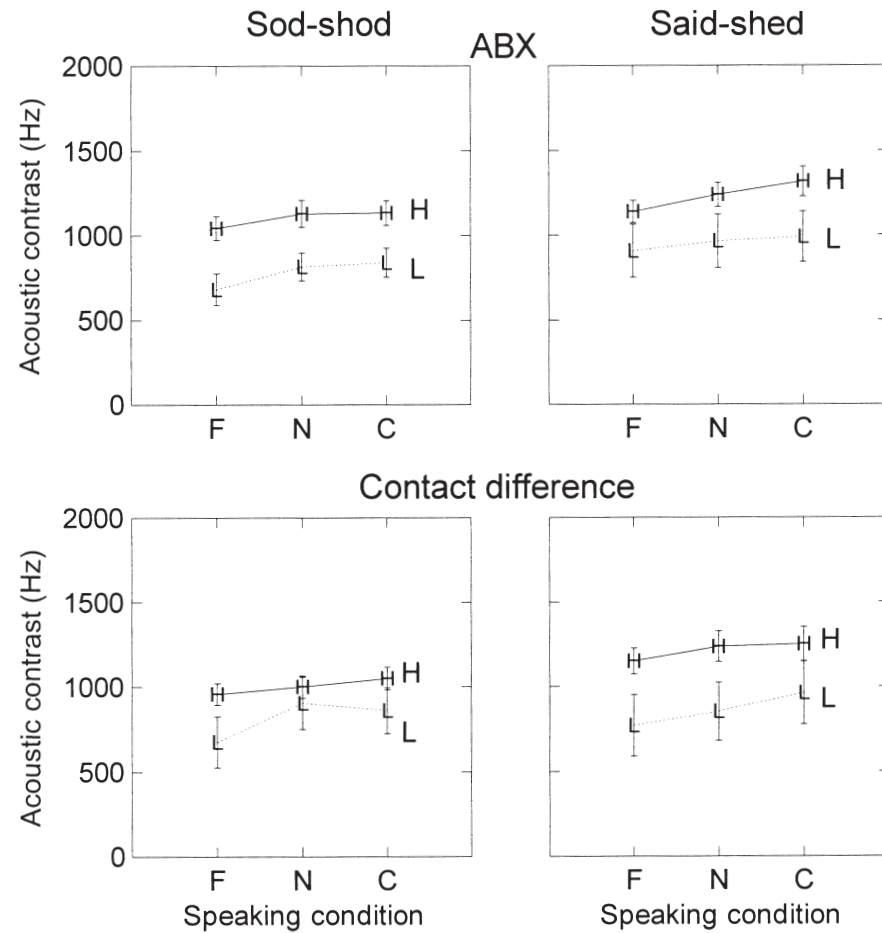
Figure 1. An example of a saturation effect is illustrated in panel A, where an abrupt change in sublingual cavity volume results from a small change in forward tongue movement. Schematized illustrations of /ʃ/ and /s/ articulation are shown in panels B and C, respectively.



Electrode used to measure amount of contact between tongue and lower incisors during /s/ and /ʃ/ production.

Subjects divided according to **High** and **Low** contact **difference** between /s/ and /ʃ/

Figure 5. The effect on sibilant acoustic contrast of two binary speaker characteristics: high (H) and low (L) use of differential contact in producing the sibilants (lower panel) and high and low acuity on an ABX test of sibilant discrimination (upper panel). Results are plotted in the left-hand column for *sod-shod* and in the right for *said-shed*. Each point is the mean of values of acoustic contrast between /s/ and /ʃ/, where the number of values represented in each mean is equal to the number of participants in the high and low groups.



Linking auditory acuity and speech production (3)

Improved discrimination test: Much finer continuum; variable step size

Perceptual acuity now correlated with Average Vowel Space (rather than a specific vowel contrast):

$$r \approx \mathbf{0.45}$$

Also correlated with average **dispersion** around category means:

$$r \approx \mathbf{-0.55}$$

Perkell, J., Lane, H., Ghosh, S., Matthies, M., Tiede, M., Guenther, F., Ménard, L. (2008). Mechanisms of Vowel Production: Auditory Goals and Speaker Acuity. Proc. 8th International Seminar on Speech Production, 29-32.

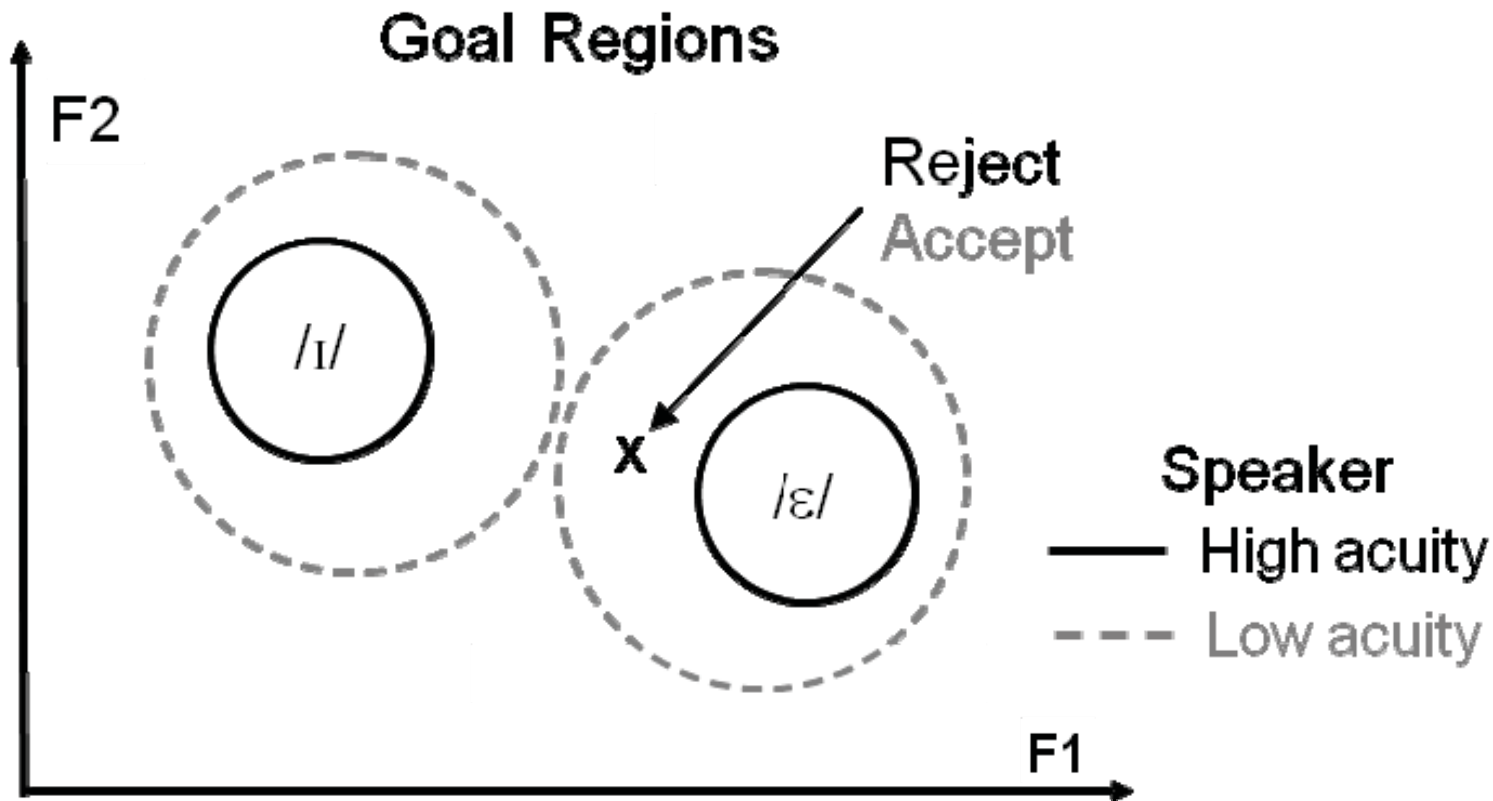


Figure 4. *Schematic illustration of goal regions of a high-acuity speaker and a low-acuity speaker for the vowels /i/ and /ε/ in F1 x F2 space.*

Linking somatosensory acuity and speech production (1)

Similar to original Perkell et al. fricative experiment.

But

- improved test of auditory discrimination

- improved spectral measurements of fricative contrast

- explicit measurement of somatosensory acuity**

Ghosh, S., Matthies, M., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F., Lane, H., Perkell, J. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity, *Journal of the Acoustical Society of America*, 128(5), 3079-3087.

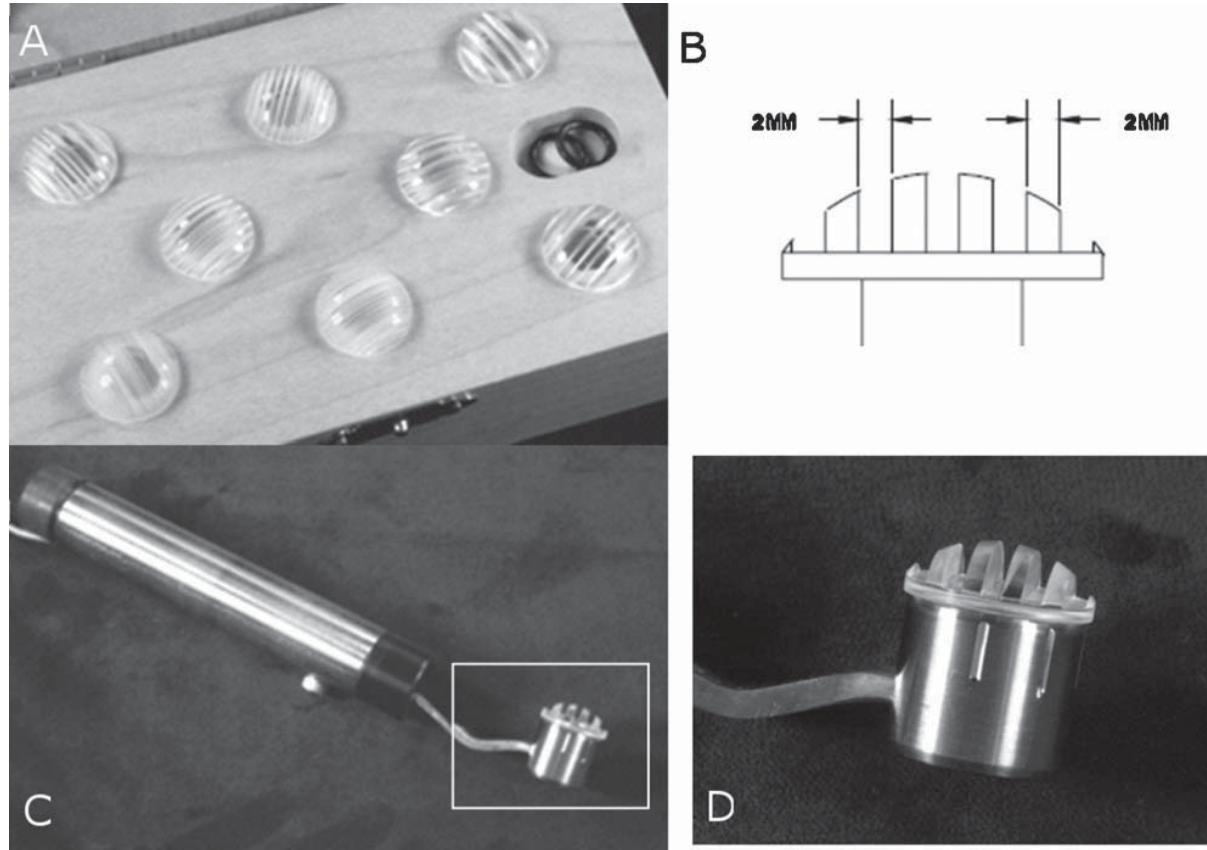


FIG. 1. JVP Domes and custom probe. (a) The set of 8 domes used in the study. (b) Grid spacing on one dome. (c) Custom holder used to apply pressure. The strain gauges are inside the handle and are not visible. (d) Magnified image of the region marked by the white box in (c).

Subject's task: Discriminate groove orientation

Advantages

Basically same procedure as for auditory acuity possible
Better than traditional two-point discrimination tests

Disadvantage

Clearly non-speech!

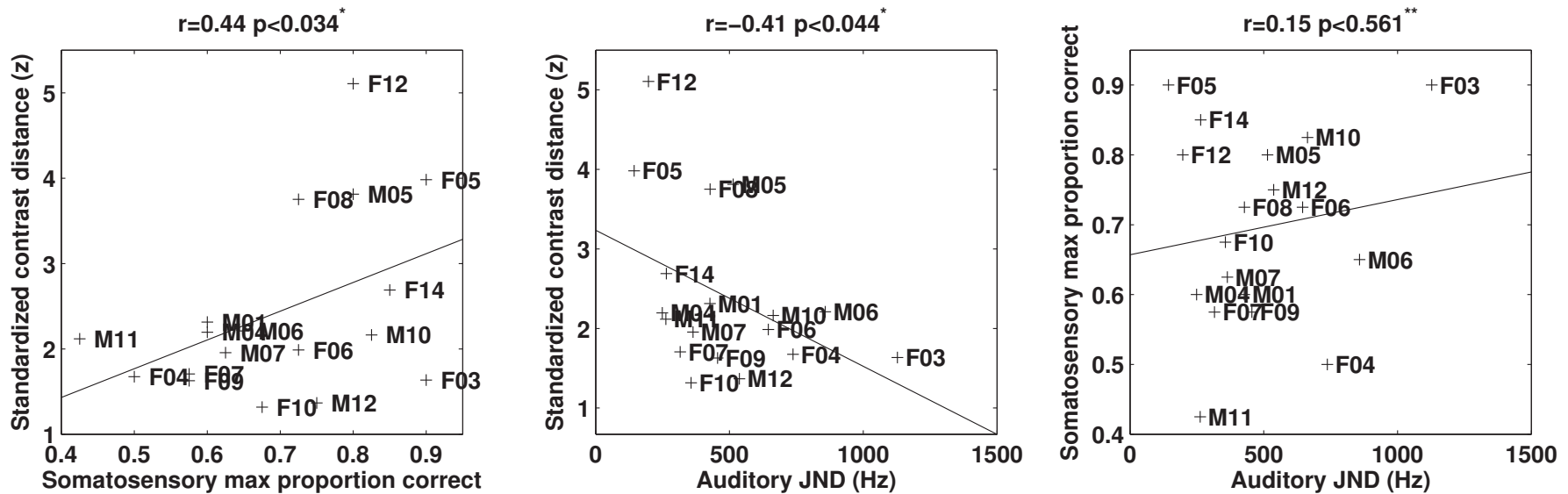


FIG. 4. Correlations between standardized contrast distance and somatosensory maximum proportion correct (left), between standardized contrast distance and auditory JND (middle) and between somatosensory maximum proportion correct and auditory JND (right). (*-one-tailed, **-two-tailed).

Left and middle panels:

Roughly same correlation between acuity in each sensory modality and the production measure (contrast distance)

Right panel:

No correlation between auditory and somatosensory acuity

Subjects categorized into High and Low discrimination performance for audition and somatosensation

Best production performance for subjects with high acuity in both modalities = group **b**

Worst performance (group **n**) with low acuity in both modalities

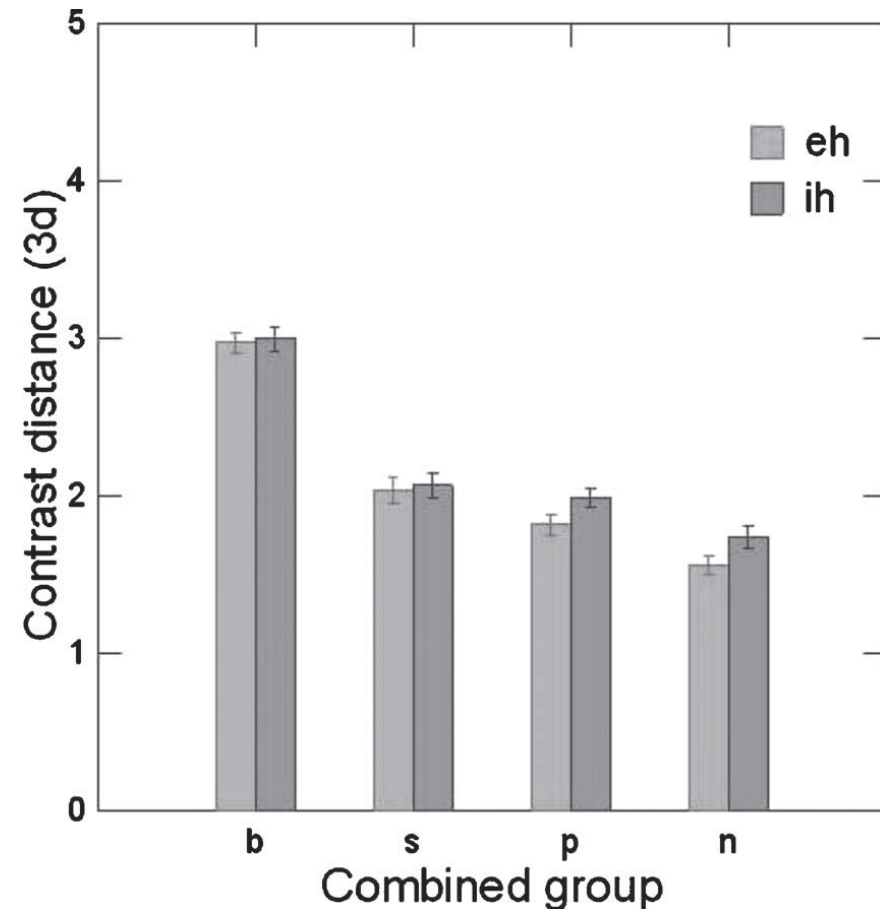


FIG. 5. Differences in 3-D contrast distance for said/shed and sid/shid as a function of group based on acuity and vowels ('eh' as in said, 'ih' as in sid; Groups-b: SomGrp=1, AudGrp=1; s: SomGrp=1, AudGrp=0; p: SomGrp=0, AudGrp=1; n: SomGrp=0, AudGrp=0; 1 indicates higher than median, and 0 indicates lower than median).

Confirmation that the grooved dome method is in the right ball-park for the sensory acuity of the articulatory organs:

Discrimination threshold roughly doubles in old subjects
Spatiotemporal stability declines in old subjects

Wohlert, A. (1996). Tactile perception of spatial stimuli on the lip surface by young and older adults, *Journal of Speech, Language and Hearing Research*, 39, 1191-1198.

Wohlert, Amy B., Smith, Anne (1998). Spatiotemporal stability of lip movements in older adult speakers, *Journal of Speech, Language & Hearing Research*, 41 (1)

More direct evidence that somatosensory representations are important in speech

Nasir, S., Ostry, D. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11(10), 1217-1222.

Tremblay, Stéphanie; Shiller, Douglas M; Ostry, David J. (2003). "Somatosensory basis of speech production", *Nature*, Volume 423, Issue 6942, 866-869

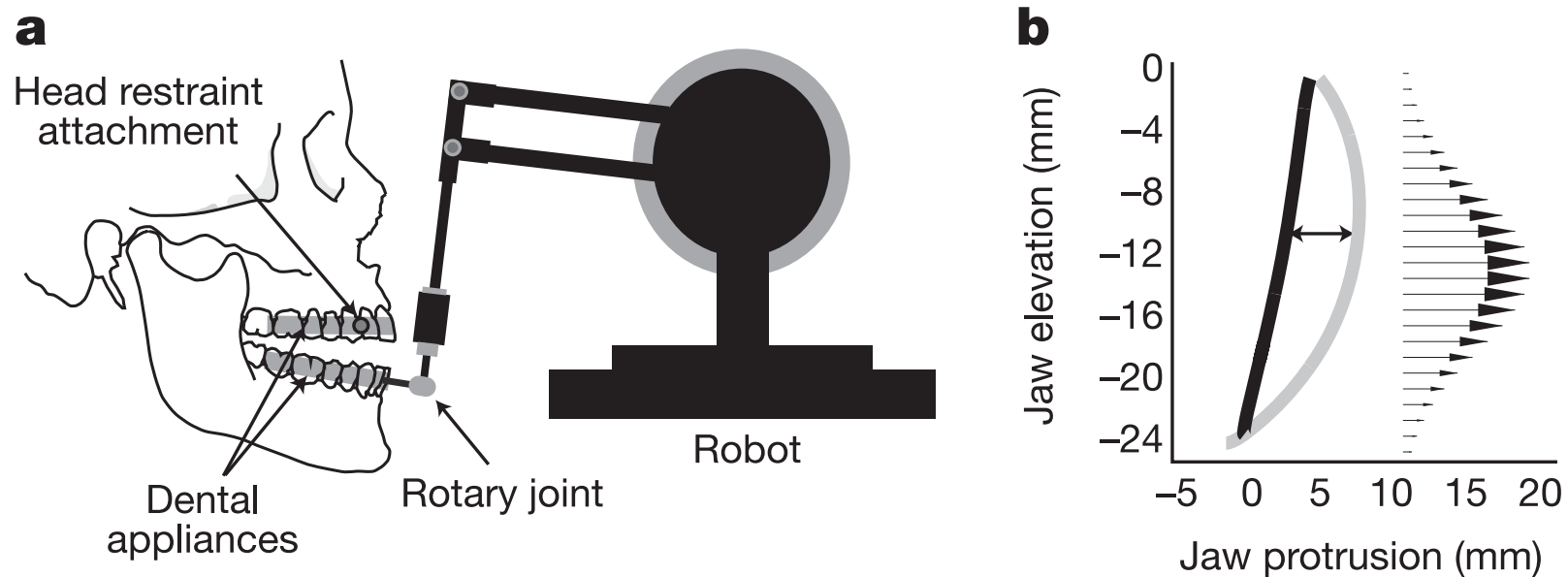


Figure 1 Experimental set-up and representative data. **a**, Diagram showing subject attached to the robotic device. **b**, Jaw opening movement with the force field off (black) and on initial exposure to the field (grey). Vectors depict the magnitude and direction of force applied by the robot over the course of the movement. The double-headed arrow shows the maximum horizontal deviation between null-field and force-field movements that served as a performance index.

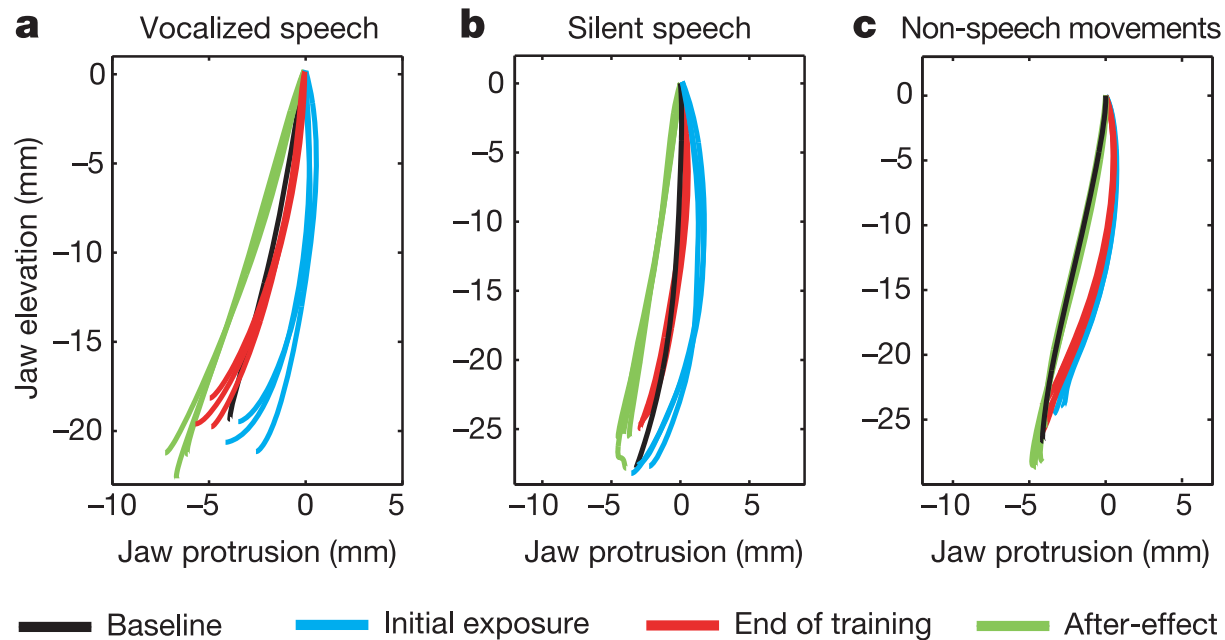


Figure 2 Sagittal plane jaw motion paths. Data were acquired during the baseline condition (black trace), on initial exposure to the force field (blue), at the end of training (red), and following unexpected removal of the field (green). The figure shows individual trials for single subjects. **a**, During vocalized speech, adaptation to the force field and a subsequent after-effect are observed. **b**, During silent speech, the pattern of adaptation and after-effect observed in vocalized speech are unaltered by removal of acoustic feedback. **c**, Matched non-speech movements show neither adaptation nor an after-effect.

Subjects must have a very precise somatosensory representation of speech movements.

Auditory information cannot be driving the compensation:
The perturbation has very little effect on speech acoustics (deaf subjects cannot hear anything anyway)

May help explain why speech deteriorates relatively slowly with prolonged deafness

Indicates speech motor learning not dependent on auditory information

Subjects probably vary in the relative weight they attach to auditory and somatosensory information

Because the 'robot' experiments are so much fun:

Evidence that manipulating the articulators influences perception:

Ito T, Tiede M, Ostry DJ (2009)
Somatosensory function in speech perception, Proceedings of the National Academy of Sciences, 106: 1245-1248

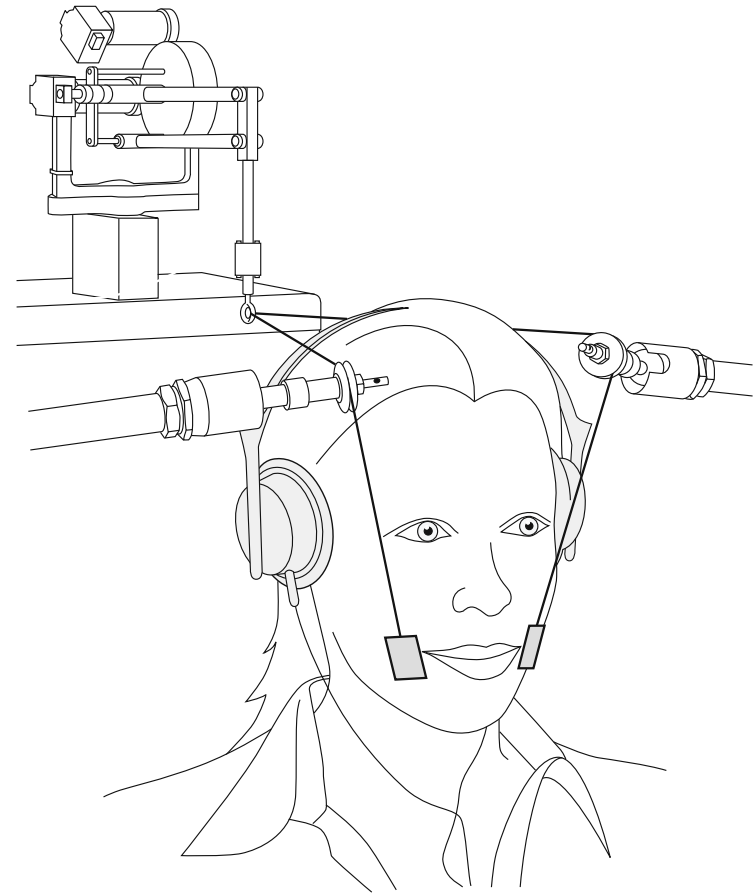


Fig. 1. Experimental setup for the delivery of skin stretch perturbations.

A possible alternative approach to testing somatosensory sensitivity in a speech-related context:

Loucks TM, De Nil LF (2006) Anomalous sensorimotor integration in adults who stutter: a tendon vibration study. *Neuroscience Letters* 402(1-2): 195-200.

The DIVA model of Frank Guenther can explain the link between auditory acuity and production precision very easily

including many findings in deaf subjects after receiving a cochlear implant

(and probably also the link with somatosensory acuity).

But there is another main source of sensory information about speech

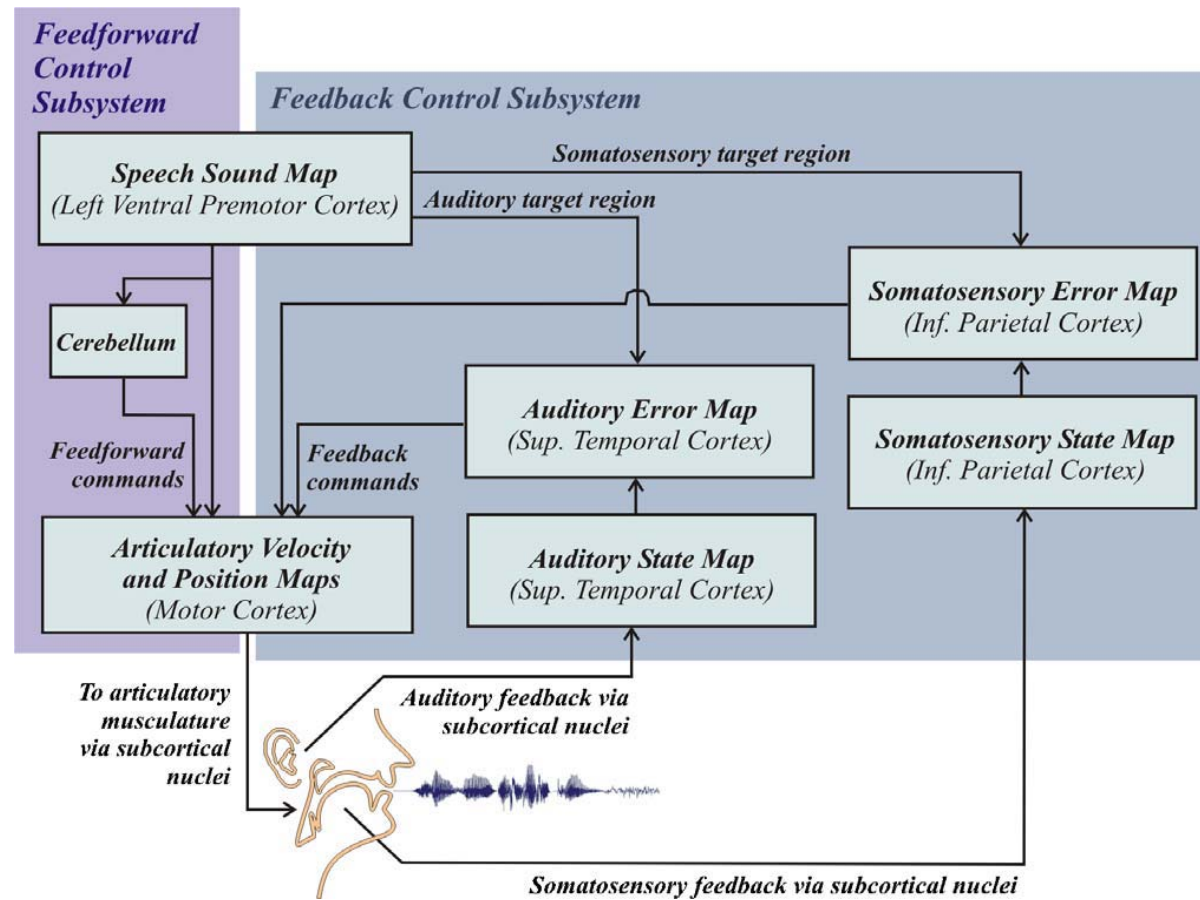


Fig. 1. Hypothesized neural processing stages involved in speech acquisition and production according to the DIVA model. Projections to and from the cerebellum are simplified for clarity.

Linking visual acuity and speech production (1)

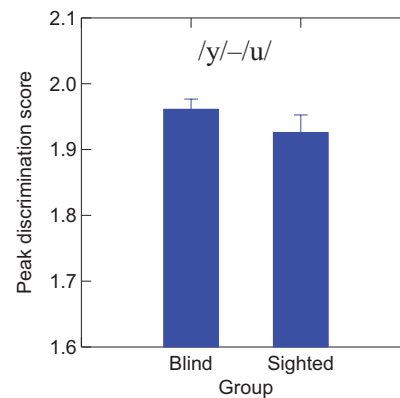
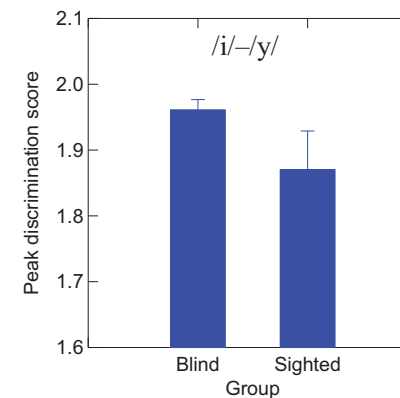
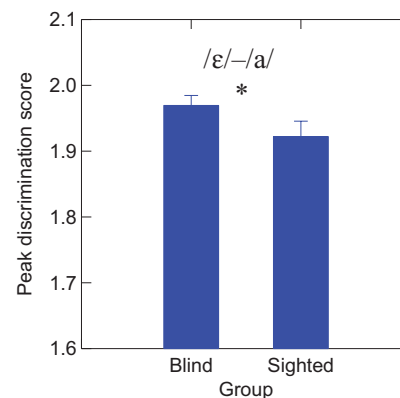
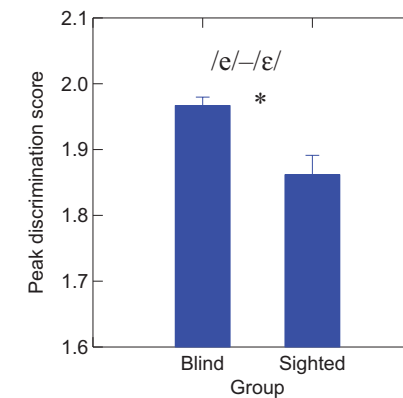
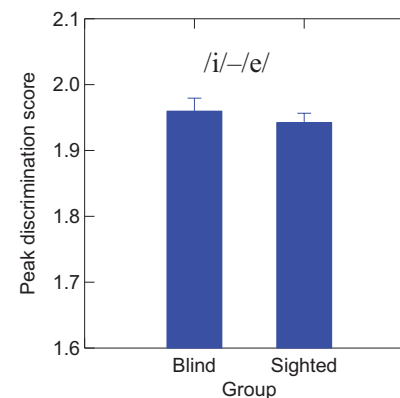
Ménard, L., Dupont, S., Baum, S., Aubin, J. (2009) Production and perception of French vowels by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America*, 126(3), 1406-1414.

Basically use the same paradigm as the earlier MIT studies to assess acuity in perception and contrast in production.

Auditory acuity (peak discrimination score) for various vowel continua.

No surprise: Overall better acuity in blind subjects

(In each panel:
left = blind
right = sighted)



Contrast in production:

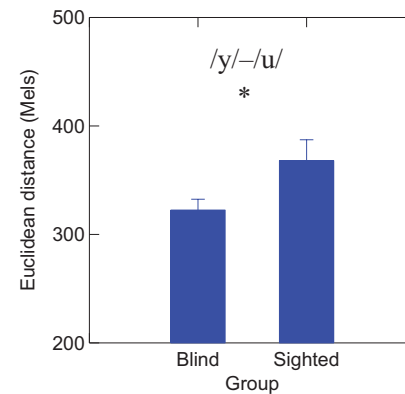
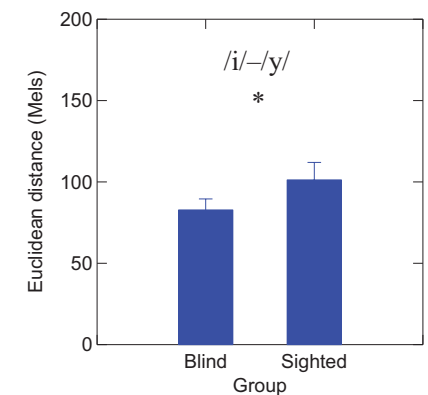
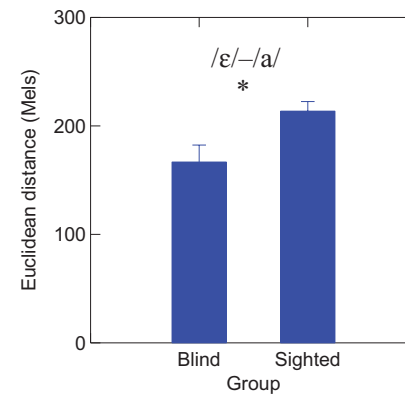
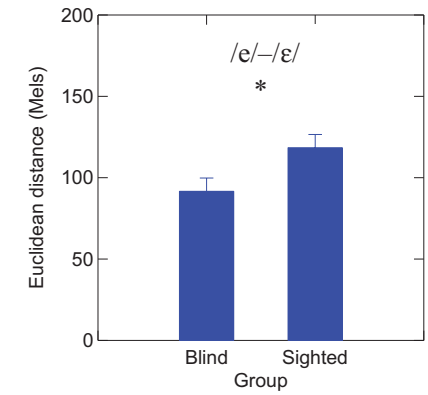
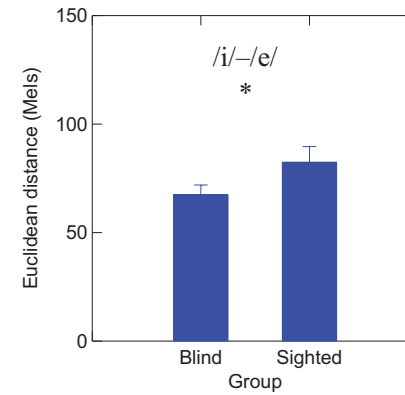
sighted subjects better!

Fowler-style explanation?

(In each panel:

left = blind

right = sighted)



Usher syndrome

So does visual sensory information help to stabilize the production system?

====>a logical extension to the acuity paradigm?

Test visual discrimination abilities for speech stimuli?

(not clear that this has been done before)

Will visual acuity also turn out to correlate with measures of contrast in production?

Linking visual acuity and speech production (2)

Further evidence from cochlear implant subjects that multisensory integration is crucial to speech.

Lachs, L., Pisoni, D., Kirk, K. (2001). Use of Audiovisual Information in Speech Perception by Prelingually Deaf Children with Cochlear Implants: A First Report. *Ear & Hearing*, 22(2), 236-251.

Bergeson, T., Pisoni, D., Davis, R. (2005). Development of Audiovisual Comprehension Skills in Prelingually Deaf Children With Cochlear Implants. *Ear & Hearing*, 26, 149-164.

Audiovisual results (1)

Following provision with the implant, measures of speech perception ability improve in **visual-only** conditions too.

“This suggests that a cochlear implant, aside from providing access to the auditory sensory modality, also allows for the construction of more general, more fully specified phonetic representations of the articulatory form of speech” (Lachs et al., 2001)

Audiovisual results (2)

Children with high values for **audiovisual gain** (= perceptual gain from adding visual to auditory information) also perform better in **auditory-only** tests of speech perception

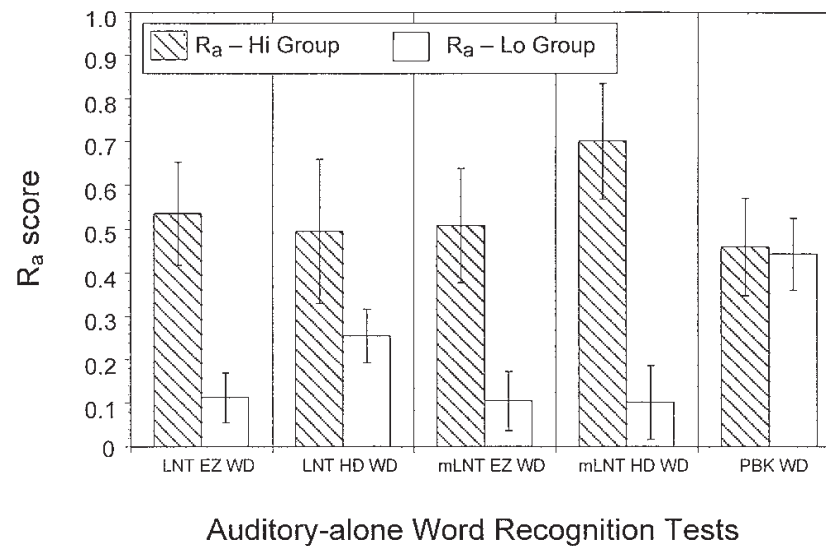


Figure 2. Audiovisual benefit (R_a) for the high-performing (above the median) and low-performing (below the median) groups of the median splits for five measures of auditory-alone word recognition. Error bars are standard errors.

This indicates that the ability to integrate multimodal information is related to the ability to use unimodal information as effectively as possible.

Audiovisual results (3)

Lachs et al. also found a correlation between audiovisual gain and speech *production* (intelligibility)

Audiovisual results (4)

Pre-implant performance in **visual-only** tests of speech perception (“lip-reading”) correlates strongly with **post-implant** performance in **auditory-only** conditions (as well as with measures of speech production)

Audiovisual results (5)

Consistent result in numerous studies of speech with cochlear implant: Children trained with the *oral communication* method perform better than those trained with the *total communication* method.

OC-children: Consistently confronted with speech as coherent auditory and visual information

TC-children: Part of their attention is distracted away from visual speech to auxiliary manual movements

Some bottom lines

“It is possible that deaf children who are most efficient at making use of any source of sensory information about speech are able to use the only available cues before cochlear implantation, that is, lipreading cues, and then once their hearing is restored via a cochlear implant, they are able to make use of both auditory and visual cues” (Bergeson et al., 2005)

“Over time, processes of perceptual learning will come to exploit lawful co-occurrences in disparate modalities, until a rich and highly redundant multimodal representation of speech emerges” (Lachs et al., 2001)

Sensory integration: Not just a question of signal-to-noise ratio

The speech performance of cochlear implant subjects is highly variable.

Why?

Our hypothesis:

When the auditory system is usable, but substandard, then speech performance depends crucially on the ability to integrate ***all*** sources of sensory information: auditory, somatosensory, visual