

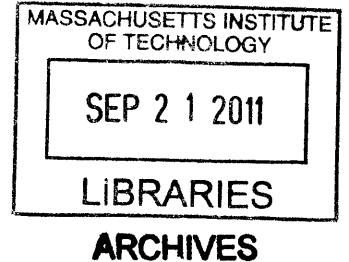
Articulatory Feature Encoding and Sensorimotor Training for Tactually Supplemented Speech Reception by the Hearing-Impaired

by

Theodore M. Moallem

B.A. Neuroscience and Behavior
Columbia College, Columbia University, New York (1998)

M.S. Neuroscience
University of California, San Francisco (2001)



SUBMITTED TO
THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2011

©2011 Theodore M. Moallem. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Handwritten signature of Theodore M. Moallem in black ink.

Signature of Author

Harvard-MIT Division of Health Sciences and Technology
September 6, 2011

Certified by

Louis D. Braidia, Ph.D.
Henry Ellis Warren Professor of Electrical Engineering and Health Sciences and Technology, MIT
Thesis Supervisor

Accepted by

Ram Sasisekharan, Ph.D.
Director, Harvard-MIT Division of Health Sciences and Technology
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering, MIT

Articulatory Feature Encoding and Sensorimotor Training for Tactually Supplemented Speech Reception by the Hearing-Impaired

by

Theodore M. Moallem

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on September 6, 2011 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Abstract

This thesis builds on previous efforts to develop tactile speech-reception aids for the hearing-impaired. Whereas conventional hearing aids mainly amplify acoustic signals, tactile speech aids convert acoustic information into a form perceptible via the sense of touch. By facilitating visual speechreading and providing sensory feedback for vocal control, tactile speech aids may substantially enhance speech communication abilities in the absence of useful hearing.

Research for this thesis consisted of several lines of work. First, tactual detection and temporal order discrimination by congenitally deaf adults were examined, in order to assess the practicability of encoding acoustic speech information as temporal relationships among tactual stimuli. Temporal resolution among most congenitally deaf subjects was deemed adequate for reception of tactually-encoded speech cues. Tactual offset-order discrimination thresholds substantially exceeded those measured for onset-order, underscoring fundamental differences between stimulus masking dynamics in the somatosensory and auditory systems.

Next, a tactual speech transduction scheme was designed with the aim of extending the amount of articulatory information conveyed by an earlier vocoder-type tactile speech display strategy. The novel transduction scheme derives relative amplitude cues from three frequency-filtered speech bands, preserving the cross-channel timing information required for consonant voicing discriminations, while retaining low-frequency modulations that distinguish voiced and aperiodic signal components. Additionally, a sensorimotor training approach ("directed babbling") was developed with the goal of facilitating tactile speech acquisition through frequent vocal imitation of visuo-tactile speech stimuli and attention to tactual feedback from one's own vocalizations.

A final study evaluated the utility of the tactile speech display in resolving ambiguities among visually presented consonants, following either standard or enhanced sensorimotor training. Profoundly deaf and normal-hearing participants trained to exploit tactually-presented acoustic information in conjunction with visual speechreading to facilitate consonant identification in the absence of semantic context. Results indicate that the present transduction scheme can enhance

reception of consonant manner and voicing information and facilitate identification of syllable-initial and syllable-final consonants. The sensorimotor training strategy proved selectively advantageous for subjects demonstrating more gradual tactual speech acquisition.

Simple, low-cost tactile devices may prove suitable for widespread distribution in developing countries, where hearing aids and cochlear implants remain unaffordable for most severely and profoundly deaf individuals. They have the potential to enhance verbal communication with minimal need for clinical intervention.

Thesis Supervisor: Louis D. Braid, Ph.D.

Title: Henry Ellis Warren Professor of Electrical Engineering and Health Sciences & Technology

Acknowledgments

Louis Braidà (thesis supervisor)

Nathaniel Durlach (thesis committee member)

Frank Guenther (thesis committee member)

Joseph Perkell (thesis committee chair)

Lorraine Delhorne

Jay Desloge

Nikolas Francis

Twenty-one anonymous experimental subjects

MIT Electronics Research Society (MITERS)

Lamine Toure, Patricia Tang, and RAMBAX-MIT Senegalese Drumming Ensemble

The author was much obliged to Charlotte Reed for advising on the study described in Chapter 3.

Dedications

To Andrew Moallem, Serena Moallem, and Dr. Anita Moallem

To all of my friends and family

In loving memory of Dr. Sha Moallem (September 1, 1933 - August 9, 2007)

May he rest in peace.

Table of Contents

Abstract	3
Acknowledgments	5
Table of Contents	7
Chapter 1_ Introduction	9
1.1 Thesis Overview	9
1.2 Hearing Impairment	9
1.2.1 Characterization of Hearing Loss	9
1.2.2 Sensory Aids and Protheses	10
1.2.3 Oral and Manual Communication Strategies	12
1.3 Speech.....	13
1.3.1 Acoustic Phonetics.....	13
1.3.2 Articulatory Phonetics	14
1.3.3 Lipreading: Visual Speech Reception	15
1.3.4 Speech Perception and Sensorimotor Representation	16
1.4 Tactual Perception.....	19
1.4.1 Tactual Sensation and Peripheral Physiology.....	19
1.4.2 Tactual Discrimination of Intensity and Frequency	22
1.4.3 Tactual Temporal Order Discrimination	24
Chapter 2_ Tactile Speech Communication: Principles and Practice	27
2.1 Natural Methods of Tactual Speech Communication	27
2.1.1 The Tadoma Method of Tactual Speechreading	27
2.1.2 The Tactiling Method.....	31
2.2 Artificial Tactual Speech Communication	32
2.3 Tactual Speech as a Supplement to Lipreading	35
Chapter 3_ Tactual Detection and Temporal Order Resolution in Congenitally Deaf Adults	37
3.1 Overview	37
3.2 Methods	38
3.2.1 Subjects	38
3.2.2 Apparatus.....	39
3.2.3 Stimuli and Procedures	40
3.2.4 Data analysis.....	44
3.3 Results	47
3.3.1 Detection thresholds	47
3.3.2 Temporal onset-order discrimination.....	48
3.3.3 Temporal offset-order discrimination.....	52
3.3.4 Effects of relative level and duration on temporal resolution.....	54

3.4 Discussion	56
3.4.1 Tactual detection	56
3.4.2 Tactual temporal order discrimination	56
3.4.3 Amplitude and duration effects	59
3.4.4 Implications for tactile displays of speech	61
3.5 Summary	63
 Chapter 4_ Development and Implementation of a Real-time Tactual Speech Display	 65
4.1 A Tactual Cue for Voicing	65
4.2 Practical Objectives and Constraints	66
4.2.1 Fine Temporal and Gross Spectral Enhancement	67
4.2.2 Suitability for Low-Cost Mobile Implementation	68
4.3 Tactual Apparatus	68
4.4 Tactual Speech Transduction Scheme	69
4.5 Vibrotactile Encoding of Articulatory Features	75
4.5.1 Vowels	75
4.5.2 Syllable-Initial Consonants	76
4.5.3 Syllable-Final Consonants	80
4.6 Speech Processing Software	83
 Chapter 5_ Tactile Speech Training and Evaluation	 85
5.1 Adapting the Tadoma Lesson Plan	85
5.2 Methods	87
5.2.1 Subjects	87
5.2.2 Tactual Psychophysics	88
5.2.3 Speech Stimuli	90
5.2.4 Training and Evaluation Protocol	90
5.3 Results	96
5.3.1 Tactual Sensitivity	96
5.3.2 Tactual Onset-Order Discrimination	98
5.3.3 Pairwise Consonant Discrimination	99
5.3.4 Twelve-Consonant Identification	105
5.3.5 Correlation of Psychophysical Measurements with Tactual Speech Reception	124
5.4 Discussion	128
5.4.1 Tactual Sensitivity	128
5.4.2 Tactual Onset-Order Discrimination	129
5.4.3 A Visual Cue for Final Consonant Voicing	133
5.4.4 Tactually-Assisted Consonant Recognition	134
5.4.5 Performance Variability among Subjects	139
5.4.6 Consideration of Sensorimotor Training Strategy	140
5.5 Summary	144
5.6 Concluding Remarks	147
 References	 150

Chapter 1.

Introduction

1.1 Thesis Overview

This dissertation is divided into five chapters. Chapter 1 provides the reader with background information concerning hearing impairment speech production and perception, tactual function and physiology.

Chapter 2 reviews previous implementations of tactual speech reception, beginning with the Tadoma speechreading method and continuing with an overview of artificial methods of tactual speech reception.

The remaining chapters describe original research, carried out in support of this thesis. Chapter 3 presents a preliminary study evaluating tactual detection and temporal order discrimination among nine profoundly deaf adults and five normal hearing adult controls. Chapter 4 describes the design and implementation of a tactual speech transduction scheme, intended to convey acoustic speech cues that supplement visual speechreading. Finally, Chapter 5 describes the training and evaluation of tactual speech reception in three normal-hearing and five profoundly deaf adults, assessing the efficacy of the speech transduction scheme and the sensorimotor training strategy developed in this thesis.

1.2 Hearing Impairment

1.2.1 Characterization of Hearing Loss

Clinically, hearing impairment is commonly classified by degree as mild, moderate, severe, or profound. Each ear is assessed separately. Asymmetric patterns of hearing loss and differing patterns of threshold elevation across frequencies can make objective classification somewhat challenging. Audiologists typically assess hearing status by administering an interactive, adaptive-level, subjective response test, in which patients indicate their detection of pure tones, ranging in frequency from 125 Hz (or 250 Hz) to 8000 Hz. By common practice, detection thresholds falling close to the normative threshold for a given frequency are classified as normal, whereas detection thresholds exceeding the normative value by more than 20-25 dB at a given frequency is regarded as indicative of hearing loss. Generally speaking, hearing loss in the range of 26-40 dB in the better ear is considered "mild" hearing loss. Hearing losses averaging 41-55 dB or 56-70 dB are referred to as "moderate" or "moderately severe", respectively. Between about 71-90 dB, hearing loss is classified as "severe", and losses beyond 90 dB are classified as "profound" (Clark,

1981). The correspondence between descriptive terminology and extent of hearing loss varies somewhat in practice.

Hearing loss is also classified with respect to the level of auditory system at which damage, occlusion, or other dysfunction manifests. Conductive hearing losses result from disruption of sound transmission through the outer ear and/or middle ear. Conductive losses can result from build-up of cerumen ("ear wax") in the ear canal, damage to the tympanic membrane, or any number of factors that interfere with middle ear transmission, including accumulation of fluids (often secondary to infection), abnormal tissue growth, and damage to ossicles (the small bones of the middle ear). Sensorineural hearing loss (SNHL) results from damage to either the cochlea in the inner ear or the auditory nerve, connecting the cochlea to the brain. In particular, the hair cells of the cochlea are susceptible to damage through exposure to loud noise, ototoxic drugs, and hair cell dysfunction is common in aging-related hearing loss (presbycusis). Hearing loss arising from damage to the nervous system central to the cochlea and peripheral auditory nerve are grouped generally as central or neural hearing losses. The manifestation of such losses can vary substantially, depending on the level of the CNS affected as well as the type and extent of damage or malformation. Of course, hearing loss can also arise due to some combination of conductive, sensorineural, and/or central factors, in which cases it is clinically characterized as a "mixed" hearing loss (Valente et al., 2010).

1.2.2 Sensory Aids and Protheses

A hearing aid is an electroacoustic device, which selectively amplifies sound for the benefit of a hearing-impaired user. These devices typically include a microphone, receiver (speaker), processing circuitry, and batteries, but the specific configuration, circuitry, and adjustability can vary considerably. Modern commercial hearing aids are most often either behind-the-ear (BTE) models, in which the bulk of the hardware sits behind the pinna and sound is fed into the ear canal through a small tube, or in-the-ear (ITE) devices, some of which protrude out into the concha and others of which are small enough to fit completely inside the ear canal. Larger, body-worn hearing aids are still in use in some early education programs and other instances in which considerations such as device power, durability, and battery life outweigh those of convenience and fashionability. The benefit imparted by hearing aids depends largely on one's level of hearing loss. For individuals with mild or moderate hearing loss, hearing aids may support relatively normal verbal communication, whereas those with bilateral profound hearing loss receive little benefit by comparison. Hearing aids equipped with telecoils can receive sound transmitted electromagnetically from certain telephones and specialized "induction loop" systems (Dillon, 2001).

Cochlear implants and auditory brainstem implants are sensory prostheses, generally restricted to individuals deemed to have little or no useful residual hearing. They consist of two distinct hardware elements. The surgically implanted portion of each device includes electrodes, designed to stimulate the auditory nervous system directly. The externally-worn portion receives acoustic signals, performs the required signal processing, and wirelessly transmits a patterned signal to a subcutaneous module that directs electrical stimulation via implanted electrodes. Cochlear implants are designed to stimulate the auditory nerve fibers that tonotopically innervate the cochlea of the inner ear. Auditory brainstem implants are designed to stimulate the auditory brainstem directly, bypassing a dysfunctional (or absent) auditory nerve (Rauschecker and Shannon, 2002).

Recent data indicate that a significant fraction of adult cochlear implant recipients can achieve relatively high levels of acoustic speech reception under favorable listening conditions (e.g., see Tyler et al., 1996; Bassim et al., 2005; Wilson and Dorman, 2008). Furthermore, studies indicate that the most successful child implantees may approach the performance levels of their normal-hearing peers in the areas of speech reception, speech production, and linguistic development (e.g., Ouellet and Cohen, 1999; Uchanski and Geers, 2003; Nicholas and Geers, 2007). However, not all deaf persons are able to be implanted or to achieve benefits from implantation. The benefits of implantation have thus far been limited to post-lingually deaf adults and those implanted as young children. Adults who were deaf prior to language acquisition usually do not develop speech reception abilities following implantation and are thus considered poor candidates for the procedure.

Tactual aids are another class of sensory aids that typically perform audio-tactual transduction, converting input from an acoustic microphone into tactually-perceptible patterns of vibrotactile or electrocutaneous stimulation. Tactual aids can vary substantially in the strategy used to encode sound for presentation to the skin. The perceived output depends on many design factors, including signal processing scheme, choice of mechanical or electrical stimulation, as well as the number, sizes, and locations of stimulator contacts on the user's body. The benefits achieved are quite variable, ranging from a basic awareness of environmental sounds to substantial enhancement of speech reception. Unlike hearing aids, tactual aids need not be precisely fitted and adjusted by a highly-trained clinician. Provided that perceptual training software is implemented on standard personal computer platforms, tactile aids can cost downwards of an order of magnitude less than hearing aids and two orders of magnitude less than cochlear implants. The next chapter discusses tactual aid development in further detail, focusing on those devices intended to benefit speech reception.

1.2.3 Oral and Manual Communication Strategies

For hearing-impaired individuals, preferred, optimal, and chosen modes of communication may be one and the same, or they may differ substantially. One's preferred communication mode can depend on many factors, including manner and temporal progression of hearing loss, the communication mode or modes used in childhood, as well as that most convenient in one's current community and daily activities. The mode of communication chosen at any given moment depends on one's own preferences and abilities as well as the preferred communication method of fellow interlocutors.

Many hearing-impaired individuals rely primarily on oral communication methods. Individuals with mild and moderate hearing impairments may benefit sufficiently from hearing aids as to allow for auditory-alone speech reception. More typically, effective communication relies heavily on visual speechreading (lipreading), either alone or in combination with one or more sensory aids (Dillon, 2001). This is especially true under low signal-to-noise ratio (SNR) conditions, which are commonly encountered in public spaces and social functions.

Manual communication methods involve production of hand and arm gestures and movements, but can also include movements and gestures involving other body parts, facial expression, or the body as a whole (e.g., posture). Signed versions of spoken languages (e.g., signed exact English, or SEE) follow the syntactic structure of the corresponding spoken language, replacing word-sounds with word-gestures. By contrast, natural sign languages (e.g., American Sign Language; ASL) are complete, self-contained languages with fully-defined gestural grammars. Thus, whereas conversion between spoken English and SEE is straightforward, conversion between spoken English and ASL requires translation and is more likely to give rise to semantic confusions (Bornstein, 1990).

Cued speech is distinct from sign languages, but rather involves supplementing verbal speech with a set of hand-shapes and movements, which serve to eliminate certain ambiguities that are inherent to visual speechreading. Thus, for example, although the consonants /p/ and /b/ may appear identical on the lips of a speaker, pairing of the utterance with one raised finger indicates that the intended consonant is /p/, whereas four raised fingers would lead one to interpret the same verbal gesture as /b/ (Cornett, 1988).

When sufficiently intact, vision typically plays a significant role in both oral and manual forms of communication. However, when the use of vision is precluded, as is the case for deafblind individuals, both oral and manual communication can be achieved through direct physical contact, supported by the

sense of touch. The Tadoma method of speechreading, although no longer in common use, offers an example of how the tactual sense can enable deafblind individuals to communicate orally. Tadoma is described in detail in the next chapter. Tactile signing and finger-spelling are manual communication methods frequently used by the deafblind today. Tactile signing is an adaptation of sign language, in which a deafblind individual places his hands in contact with the hands of the signer, thereby perceiving (and inferring) hand gestures, movements, and shapes. The deafblind individual's own signing may be conveyed either visually or tactually. In tactual finger-spelling, information is conveyed character-by-character, with one individual using a finger to trace symbols onto the palm of another (Reed et al., 1985, 1990, 1995).

1.3 Speech

1.3.1 Acoustic Phonetics

Speech sound sources can be generally classified as periodic, aperiodic, or some combination of the two. Periodicity is characteristic of vocalization, produced by vibration of the glottal folds (vocal cords). The pulsation frequency of the glottis determines the fundamental frequency (F_0) of voicing and concentrates spectral energy at F_0 and its harmonics. The resulting harmonic complex is acoustically filtered by the vocal tract, which further shapes the sound spectrum. Natural frequencies of the vocal tract filter, referred to as formants, are evidenced by local peaks in the vocal spectrum. The first few formants and their dependence on changes in the vocal tract cross-sectional area function are largely responsible for the variety of speech sounds generated by the human articulatory apparatus. In the context of a particular spoken language, some of these speech sounds map onto a characteristic set of phonemes, including vowels and consonants, while other speech sounds are neglected. Phonemes are generally considered the smallest subdivisions of sound that combine to form distinct utterances. Vowels are articulated with an unstricted vocal tract and distinguished by their characteristic formant patterns. Consonants are produced by the build-up of air pressure behind a partial or full constriction along the vocal tract and distinguished by the location and completeness of the constriction.

Aperiodic speech sounds typically arise as a result of turbulent airflow at a constriction along the vocal tract or at the glottis itself. They also include transient bursts, resulting from the release of a full constriction following build-up of pressure behind it. Aperiodic speech sounds are also subject to spectral shaping by the vocal tract, which plays a key role in distinguishing consonants as well as whispered vowels. In whispered speech, the vocal tract is excited acoustically by aperiodic turbulent noise created

by forcing air through the adducted vocal cords. Similarly, fricative consonants are produced by forcing air through constrictions at various locations along the vocal tract. For the fricative consonants /s/ and /z/, turbulence arises at a constriction produced by positioning the tip of the tongue just behind the upper front teeth, whereas the turbulent noise of the fricatives /f/ and /v/ arises at a constriction between the upper lip and lower front teeth.

Prosody refers to the pattern of stress, timing, and intonation of speech, essentially the modulation of phonetic properties that imposes a cumulative semantic effect at the suprasegmental level (e.g., emphasis, relation, indifference or interrogatory intent).

1.3.2 Articulatory Phonetics

The mechanisms by which human articulatory apparatus produces speech sounds are well approximated acoustically by the "source-filter model" of speech production. At the heart of the model is an acoustic system in which the vocal tract is represented by a tube filter, closed at one end, where it is excited by a variable sound source consisting of voicing at the glottis and/or noise at different locations along the vocal tract. The filtering properties of an acoustic tube vary predictably as a function of its length and cross-sectional area along that length. The combined source-filter system output is the convolution product of the sound source signal and the filter response function. The natural (resonant) frequencies of the acoustic tube are calculable, all the more easily if a simple, step-wise area function is assumed.

As noted above, these resonant frequencies are commonly referred to as "formant frequencies" in the context of the vocal tract model. For a tube of uniform cross-sectional area, closed at one end, the formant frequencies are simply $F_n = [(2n-1)/4] \cdot (c/l)$, where n is the formant number, c is the speed of sound through the air inside the tube, and l is the tube length. The first few formant frequencies are the ones that contribute most to phoneme quality. Even without deriving a new formula for each new cross-sectional area function, it is possible to approximate the effects that specific perturbations to the area function will have on each of the first few formant frequencies. In this manner, it is possible to understand the effects of changing the relative positioning of articulators on formant structure and speech sound. The model can then be extended to account for non-zero radiation impedance beyond the open (mouth) end of the tube, coupling between the vocal tract and neighboring resonant cavities, and sound absorption by the vocal tract walls (Stevens, 1999).

English consonants are effectively characterized in terms of the articulatory features voicing, manner of articulation, and place of articulation. Constants are classified as either voiced or unvoiced/voiceless. Generally speaking, a voiced consonant (e.g., /b/, /d/, /z/, /v/) is one that is typically pronounced with vibration of the vocal cords, although it may be "devoiced" under some speaking conditions. By contrast, an unvoiced consonant (e.g., /p/, /t/, /s/, /f/) is strictly aperiodic, never accompanied by vocal vibration.

Manner of articulation (or manner of constriction) describes the roles of the various articulators (e.g., tongue, jaw, teeth, lips) in producing a consonant sound, by shaping and tensioning vocal tract structures appropriately. Plosives (or stops), fricatives, and affricates are all obstruent consonants (produced by constriction of airflow) but differ in manner of constriction. Plosives (e.g., /b/, /t/) are produced by building pressure behind a complete constriction, which is released to create a "burst". Fricatives (e.g., /z/, /f/) are produced by forcing air through a narrow constriction, thereby creating turbulent noise. Affricates (e.g., /tʃ/, /dʒ/) are plosive-fricative composites, beginning with a full stop, and then releasing into fricative turbulence. Approximants, nasals, and taps are all sonorant consonants (produced in a vowel-like manner without turbulent airflow). Approximants (e.g., /l/, /w/) are effectively time-varying vowels. Nasal consonants (e.g., /m/, /n/) are produced with the velum (soft palate) lowered, coupling the nasal cavity with the rest of the vocal tract.

Place of articulation describes the location at which articulators constrict the vocal tract during consonant production. There are many more possible places of articulation than are actually used to distinguish consonants in any one language. The consonants of English include labial consonants (e.g., bilabial: /b/, /p/; labiodental: /f/, /v/), coronal consonants (e.g., alveolar: /t/, /s/; dental: /θ/, /ð/), and dorsal consonants (e.g., velar: /k/, /g/; palatal: /j/). In other languages, consonants with more than one place of articulation are used, often referred to as doubly-articulated consonants (Ladefoged and Maddieson, 1996).

1.3.3 Lipreading: Visual Speech Reception

Lipreading (also referred to as speechreading) is a method of speech reception through a combination of visual recognition of articulatory cues and utilization of semantic context. Lipreading abilities vary substantially across individuals. Individuals with severe and profound hearing impairment may depend heavily on lipreading for verbal communication. Visually-perceived facial cues, when available, can also influence acoustic speech reception by normal-hearing listeners (McGurk and MacDonald, 1976). However, speech reception through vision alone differs fundamentally from speech reception through audition alone. Whereas a well-enunciated acoustic speech signal, presented at a sufficiently high SNR,

can effectively convey sufficient articulatory information for identification of each individual phoneme, a significant portion of those articulatory cues are usually not evident on the face of a speaker. Vibration of the vocal folds is perhaps the most ubiquitous and essential articulatory feature that is hidden from view. Visible prosodic and coarticulatory cues that do tend to correlate with vocalization (e.g., syllable duration for emphasis and in conjunction with final consonant voicing) are not consistent across speakers or semantic context. Thus, consonant voicing, vocal pitch contour, and vowel duration are poorly received through lipreading. Articulatory manner is also poorly transmitted through visible facial cues, such that visual confusions among plosive, fricative, and nasal consonants are common among lipreaders. Place of articulation can be discerned far more effectively than either voicing or manner. For example, labial, labiodental, dental, and alveolar consonants are readily distinguishable on the faces of most speakers. Phonemes articulated with the back of the tongue and pharynx produce far subtler visible cues.

To the extent that articulation is evidenced on a speaker's face, lipreading skills can be developed through perceptual training, and both hearing-impaired children and late-deafened adults stand to benefit substantially from such training. However, effective lipreading necessarily depends heavily on semantic context. Successful lipreaders enter into a verbal interaction far more attentively than one who can rely on auditory perception and auditory memory. Environmental and social cues, current events, and various preconceptions factor strongly into one's interpretation of visually-perceived speech cues, which in turn influences one's subsequent inferences (Heider and Heider, 1940; Erber, 1974; Dodd, 1977; Jeffers and Barley, 1977; Walden et al., 1977, 1981).

1.3.4 Speech Perception and Sensorimotor Representation

Several lines of evidence suggest that the capacity for speech perception is directly linked to certain aspects of speech production. In a "speech shadowing" task, listeners can verbally repeat isolated words or nonsense syllables with very short latencies, on the order of 150-250 ms (Marslen-Wilson, 1973; Porter and Lubker, 1980), indicating that an articulatory representation is achieved quite rapidly. Some listeners can shadow full spoken sentences with similarly short latencies, trailing the original speech by as little as one syllable (Marslen-Wilson, 1973). Porter and Lubker (1980) found that subjects could imitate a vowel-change (in a shadowing task) more quickly than they could respond to the same vowel-change with a predetermined vocalization. They interpreted this finding as evidence of "overlapping and intimately-linked processes of auditory analysis and speech-gesture control". In other words, the shorter response time for imitation suggests that neural representations of vowel percepts are more closely linked to the corresponding articulatory gestures (used to produce them) than to arbitrarily chosen articulatory gestures.

A number of studies have established direct correlations between an individual's perceptual tendencies and characteristic productions of certain phonetic elements. For example, Newman (2003) found subjects' pronunciation of certain phonemes to be correlated with their own "perceptual prototypes" of these phonemes, which were identified by having subjects rate the quality of phonetic stimuli that were varied along an acoustic continuum (e.g., the VOT continuum for phonemes /b/ and /p/). Similarly, Perkell et al. (2004) showed that subjects who demonstrated the greatest perceptual acuity for vowel contrasts were likely to exhibit greater acoustic contrasts in their vowel productions. Moreover, when Bradlow et al. (1997) trained native Japanese speakers in the perceptual identification of the English /r-/l/ contrast (using naturally produced /r-/l/ minimal word pairs), they found corresponding improvements in their subjects' /r/ and /l/ productions, which was reflected in the accuracy of identification by English listeners. Short-term sensory manipulations, such as "selective adaptation" of phonetic elements through repeated presentation, can also affect both speech perception and production (Eimas and Corbit, 1973; Cooper, 1979), and selectively altering a speaker's auditory feedback of his own vowels can elicit compensatory articulator activity that is sustained until normal perceptual feedback is restored (e.g., Houde and Jordan, 1998).

Motor Theories of Speech Perception

Browman and Goldstein (1986, 1992) offer a useful (albeit simplified) bridge between classical notions of speech perception and production as distinct phenomena, on the one hand, and motor theories of speech, which tend toward the opposite conclusion. Their "articulatory phonology" looks to the processes of articulation to provide an orderly, functional representation of speech sounds. Rather than theorizing as to the precise representation of speech sounds in the human CNS, articulatory phonology emphasizes that "the movement of the articulators through space over time...constitutes an organized, repeatable, linguistically relevant pattern", and that relatively complex phonological phenomena can be described relatively simply as patterns of articulatory gestures that overlap in time (Browman and Goldstein, 1986).

"To perceive an utterance...is to perceive a specific pattern of intended gestures." Such is a primary assertion of Liberman and Mattingly (1985), who along with colleagues based out of Haskins Laboratories, are most commonly associated with "the motor theory of speech perception" (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman, 1996). According to their model, coarticulation results in extensive overlapping of acoustic information corresponding to successive phonetic units in a manner that varies with phonetic context and rate of articulation (among other factors), such that there is "no straightforward correspondence in segmentation between the phonetic and acoustic representations of

the information" (Liberman and Mattingly, 1985). The same general idea, that movements of the vocal tract are effectively the objects of speech perception, is advanced by Fowler (1986) as a tenet of her "direct realist" approach to speech perception, although she and likeminded colleagues remain skeptical of motor theory's finer points (e.g., Fowler, 1996; Galantucci et al., 2006). While motor theories have not been openly embraced by a majority of speech scientists, the most basic conviction held by Liberman, Fowler, and other motor theorists is, in principle, compatible with a variety of other speech perception models (e.g., McClelland and Elman, 1986; Traunmüller, 1994; Stevens, 2002). The concept of "perceiving articulatory gestures" may seem at first glance to differ fundamentally from psychophysical notions of "perceiving speech sounds", but the two perspectives are in fact no less consistent with one another than, for example, "seeing a tree" and "seeing patterns of light reflected from a tree".

Motor theories of speech perception suggest that speech acquisition entails learning to group similar sounds that are articulated somewhat differently into phonemic categories, which helps to explain the lack of invariance between phonetic elements and their acoustic manifestations. Thus, a noise burst that is spectrally centered around 1440 Hz is most often perceived as the labial stop consonant /p/ when placed before the vowel /u/, but it is more commonly perceived as the velar stop consonant /k/ when preceding the vowel /a/ (Liberman et al, 1952). Similarly, synthesized intermediaries on the /sh/-/s/ continuum are more frequently perceived as the fricative /s/ when preceding the vowel /u/ than when preceding the vowel /a/ (Kunisaki and Fujisaki, 1977). Such perceptual distinctions are not at all surprising if one supposes that the listener can anticipate the acoustic effects of coarticulation.

The notion that speech sounds are linked to articulatory gestures in a central speech representation offers insight into findings that speech-related inputs from multiple sensory modalities can be integrated to form a unified speech percept. Such findings include the modulation of auditory speech perception by tactual (Gick & Derrick, 2009) and visual (McGurk & MacDonald, 1976) stimuli, the influence of exogenously induced facial skin stretch on the perception of speech sounds (Ito et al., 2009), and the perception of speech via tactual detection of articulatory cues by deafblind Tadoma users (Reed et. al, 1985).

In formulating their motor theory, Liberman and colleagues stress the notion that speech perception is somehow "special" and must be considered completely separately from all other perceptual phenomena (Liberman et al., 1967; Liberman and Mattingly, 1985). This dogmatic assertion may well underlie much of the contention faced from detractors of the motor theory, as well as the lack of support from others who otherwise concede the importance of articulatory gestures in the neural representation underlying speech perception. The motor theory postulates that speech perception occurs via a special "speech mode" or

"module", consisting of specialized neural structures that automate the mapping of sensory input onto a gestural framework, allowing phonetic perception that bypasses common psychoacoustic processes (Liberman and Mattingly, 1985). By analogy, auditory localization based on interaural time difference (ITD) occurs automatically via specialized processes, which are largely distinct from those processes engaged when one judges temporal ordering among acoustic stimuli. In fact, a "motor theory of sound localization" can provide substantial insight into the organization and plasticity of the neural processes underlying the localization of acoustically distinct sound sources. For example, noting that equivalent spatial displacements of two sound sources might be associated with very different acoustic changes, Aytekin et al. (2008) argue that the auditory system interprets these changes as comparable because "they can be compensated by the same set of motor actions" (namely, equivalent rotation of one's head). However, whereas the motor theory of speech perception is laden with provocative assertions such as "sounds are not the true objects of perception" and "adaptations of the motor system...took precedence in the evolution of speech" (Liberman and Mattingly, 1985), Aytekin and colleagues have merely proposed "a sensorimotor approach to sound localization". Their assertion that "sensorimotor early experience, during development, is necessary for accurate sound localization" will not likely encounter staunch opposition.

1.4 Tactual Perception

The adequacy of the tactual sense for reception of speech information may be considered from two general perspectives. On the one hand, the effectiveness among certain deafblind individuals of the Tadoma method of tactual speechreading (described in detail in the next chapter) offers evidence that tactually perceived articulatory cues can support speech comprehension. On the other hand, in developing a tactual display strategy in which speech cues are artificially encoded, the functional characteristics of the somatosensory system offer specific design constraints, which may further provide insight that facilitates troubleshooting and optimization of speech reception via tactual channels.

1.4.1 Tactual Sensation and Peripheral Physiology

A common approach to characterizing tactual perception is to measure the minimum level at which sinusoidal vibrations of a given frequency are consistently detected. Obtaining such detection threshold measurements at many frequencies provides a description of the range of tactual sensitivity at a particular site on the body. Repeating these measurements at multiple body sites, using a stimulator contact of the same size and shape, reveals that vibrotactile sensitivity functions vary substantially over the surface of the body (e.g., Verrillo, 1963, 1966, 1971). Measurements at the glabrous surface of the fingertips, which

are among the most tactually sensitive locations on the skin, indicate that maximal sensitivity to sinusoidal vibration is in the range of 200-300 Hz. At frequencies above about 80 Hz, detection threshold varies with the contact area over which stimulation is applied, such that a doubling of the contactor area results in roughly 3 dB threshold reduction (Verrillo, 1963). Moreover, for vibratory pulses shorter than about 200 ms, detection threshold increases as the stimulus duration decreases further (Verrillo, 1965).

Sensitivity falls off gradually as frequency increases above 300 Hz, such that larger displacement amplitudes are required to elicit sensation. While one cannot specify precisely the upper frequency limit of tactual sensitivity, detection thresholds at 1000 Hz are typically elevated by more than 20 dB relative to those observed at 250 Hz (Verrillo and Gescheider, 1992). Tactual sensitivity also decreases as sinusoidal frequency falls below about 200 Hz, but proper characterization is complicated somewhat by the fact that tactual sensitivity at lower frequencies is mediated by activation of several different types of mechanoreceptors in the skin. Below about 50 Hz, detection threshold is much less variable across frequencies, suggesting that a distinct set of neuronal afferent mechanisms are involved in detection (Bolanowski et al., 1988). As the frequency of sinusoidal vibration decreases from 200 Hz to less than 1 Hz, not only are larger stimulus amplitudes required to elicit a sensation, but the corresponding perceptual quality changes, and neural recordings have shown that the distribution of responsive mechanoreceptive afferents types also changes substantially. While measurement of frequency-dependent detection thresholds is relatively straightforward, connecting perceptual quality with the underlying mechanoreceptive response has proven more challenging.

Focusing on the glabrous (hairless) skin surfaces of the hands and fingers, physiological studies have identified four different types of mechanoreceptive afferent nerve fibers, characterized generally by their rate of adaptation to sustained mechanical perturbation. The two "slowly adapting" fiber types are commonly labeled SAI and SAII, and the two "rapidly adapting" fiber types are labeled RA and PC. SAI and RA afferent fibers have small receptive fields (the area of skin over which each fiber responds). Anatomical studies have linked SAI afferents with Merkel disk receptors (or Merkel cell-neurite complexes) in superficial skin layers, which respond strongly to pressure, as produced by static indentation of the skin (roughly 0.4-2.0 Hz). RA afferents have been linked with Meissner corpuscles, also located superficially in the skin, which mediate light touch, slip, and flutter sensations, responding mainly to stimulation frequencies in the range of about 2-40 Hz. SAII and PC afferent fibers have notably larger receptive fields. Anatomical studies have linked SAII afferents with Ruffini corpuscles (or Ruffini end organs) located in the subcutaneous tissue, which respond strongly to skin stretch and positional changes of the hand and fingers. Finally, PC afferents terminate in Pacinian corpuscles,

specialized nerve endings that are particularly sensitive to smooth vibrations over a very wide range of frequencies, ranging from around 40 Hz to more than 500 Hz (Greenspan and Bolanowski, 1996; Craig and Rollman, 1999).

On the basis of detailed psychophysical observations, Bolanowski et al. (1988) functionally isolated tactual perception mediated by four distinct channels at the glabrous skin. Their aim was to synthesize behavioral, physiological, and anatomical findings into a single coherent model. They described the response characteristics of the four channels, which they labeled PC, NPI, NP II, and NP III (Pacianian and non-Pacianian I-III), and they associated each channel with one of the four known mechanoreceptive afferent types. Figure 1-1 shows the threshold-frequency characteristics of the four psychophysically inferred channels, the responses of which were isolated by varying masking conditions, skin temperature, and stimulus duration. The PC channel shares the response characteristics of PC afferent fibers and is thought to respond to activation of Pacinian corpuscles, over a frequency range of roughly 40-500 Hz, which in turn produces a sensation of (smooth) "vibration". The NPI channel is physiologically mediated by RA afferent fibers in response to stimuli ranging in frequency from about 2-40 Hz, which activate Meissner corpuscles in the skin, producing a "flutter" sensation. The NP II channel is mediated by SA II afferent fibers, driven by frequencies in the range 100-500 Hz, which activate Ruffini corpuscles and produce a "buzzing" sensation. Finally, the NP III channel reflects SA I afferent fiber activity, which is driven 0.4-2.0 Hz activation of Merkel disk receptors, which gives rise to a "pressure" sensation.

Bolanowski and colleagues note that the psychophysically measured detection threshold at any frequency will most likely be determined by the mechanoreceptor channel exhibiting the lowest activation threshold. This is consistent with average detection threshold measurements from five observers, which are superimposed as large dots in Figure 1-1. This threshold would further be expected to vary as a function of such factors as skin temperature, stimulus duration, and contactor size. Moreover, their model suggests that the perceptual quality elicited by any given tactual stimulus results from the combined neural activity of all activated channels (whether it is a single channel or some combination of the four).

From a perceptual standpoint, the tactual sense is often regarded as a composite of several submodalities, roughly corresponding to the variety of percepts associated with the somatosensory nervous system. The term "vibrotactile" generally refers to sensations produced by high frequency, low-amplitude, vibratory stimulation, which is effectively transmitted along the skin surface. By contrast, the term "kinesthetic" refers to the sensation or awareness of joint angles, muscular tension, and positions or gross movements

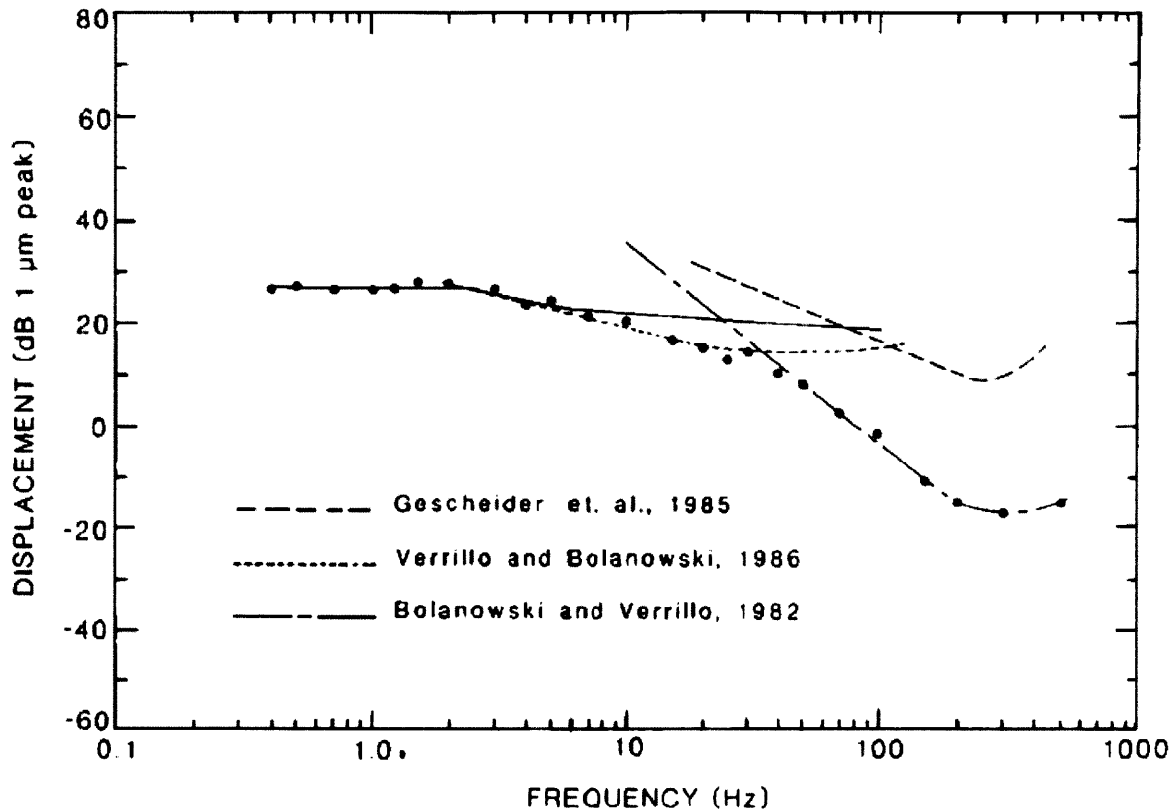


Figure 1-1. The four-channel model of vibrotactile perception showing the threshold-frequency characteristics of the various channels: — PC; ····· NP I; - - - NP II; and — NP III. Data points (large dots) superimposed on the model are average detection thresholds of five observers, with contactor size of 2.9 cm^2 and skin temperature 30°C . (Reproduced from Bolanowski et al, 1988)

of limbs and extremities. A range of other sensations, including localized pressure and "flutter", and movement on the skin might be characterized as "tactile" sensations. The term "tactual" is applied generally to all stimuli spanning the kinesthetic to vibrotactile continuum, and depending on the context, it might also encompass stimuli that produce sensations of warmth, coolness, and pain.

1.4.2 Tactual Discrimination of Intensity and Frequency

The functional intensity range of the tactual system is limited to about 55 dB above the detection threshold (i.e., 55 dB SL). Beyond this range, tactual stimuli quickly become unpleasant or painful. This dynamic range may be considered limited when compared with the auditory system, which is capable of processing stimuli with intensities up to 130 dB above detection threshold (Verrillo and Gescheider,

1992). Of course, comparing the two sensory systems in this manner may be misleading, as the 55 dB dynamic range of the tactual system applies to mechanical stimulation at a localized area on the tactual receptive surface, whereas a given acoustic stimulus tends to excite cochlear hair cells along a substantial portion of the basilar membrane.

Knudsen (1928) measured the difference limen (DL) for intensity of 64, 128, 256, and 512 Hz sinusoidal vibrations delivered to the tip of the index finger. DL was found to be relatively constant across frequencies, but it decreased with increasing absolute intensity. At 35-40 dB SL, the DL was found to approach approximately 0.4 dB. Using 60 Hz vibrotactile stimuli presented to the chest, Geldard (1957) measured DLs of 3.5 dB at 20 dB SL, which decreased to about 1.2 dB at 45 dB SL. Craig (1972) measured intensity discrimination using 200 ms bursts of 160 Hz sinusoidal vibration as well as 2-ms square wave ("tap") stimuli, both delivered to the index finger. DLs for the vibratory stimuli were about 1.5 dB, remaining relatively constant in the intensity range of 14-35 dB SL. DLs for tap stimuli were also about 1.5 dB for absolute intensities around 30 dB SL, but as the tap level was reduced to 14 dB SL, the DL rose to 2.5 dB. Finally, Gescheider et al. (1990) measured intensity discrimination using sinusoidal vibrations of 25 Hz and 250 Hz delivered to the thenar eminence of the hand. They found intensity DL to be independent of frequency, decreasing from 2.5 dB to 0.7 dB as stimulus intensity increased from 4 dB SL to 40 dB SL. Thus, as a matter of general agreement, tactual intensity discrimination is found to improve with increasing absolute intensity (as in audition).

To a greater extent than observed in the auditory system, tactual perception of frequency and intensity are intertwined, such that varying the level of a periodic vibratory stimulus will often change the subjective "pitch".

Moreover, difference limens for frequency discrimination have been found to vary substantially depending on the shape of the test waveforms (e.g., sinusoidal vs. square pulses). Knudsen (1928) reported frequency DLs in the range of 15-30% for 64, 128, 256, and 512 Hz sinusoidal vibrations presented to the index finger at 34 dB SL. Goff (1967) presented 25-200 Hz, one-second sinusoidal vibrations to the index finger. At 20 dB SL (re 100 Hz detection threshold), he measured frequency DLs in the range of 18-36%, and increasing stimulus intensities by 15 dB, he measured frequency DLs between 31-55%. Mowbray and Gebhard (1957) noted that "sine waves offer special difficulties, because frequency and intensity tend to be confounded", and they instead measured difference limens for repetition rate of mechanical pulses, at ten frequencies between 1 and 320 Hz (i.e., pulses per second), at stimulus intensities of approximately 17-26 dB SL. Pulse stimuli were presented via a 50 mm² plate held

between two fingers of one hand. Subjects switched between a standard and a comparison signal, adjusting the frequency of the comparison stimulus to match the standard. DLs for frequency were measured as the frequency difference between the two signals after adjustment, and measurements were averaged over five subjects. Frequency DLs ranged from 2% at both 1 Hz and 2.5 Hz to 8% at 160 Hz and 320 Hz. Thus, several factors influence tactual frequency DLs, including waveform shape, stimulus level, and contactor area. One should also note that the method of presentation used by Mowbray and Gebhard (plate held between two fingers) contacts multiple sites, and the user could likely vary the mechanical load applied by the fingers, thereby varying the system response.

1.4.3 Tactual Temporal Order Discrimination

Previous studies of temporal order discrimination in the tactual domain have commonly used transient, mechanical or electrocutaneous stimuli, which do not overlap in time (e.g., Hirsh and Sherrick, 1961; Sherrick, 1970; Marks et al., 1982; Shore et al., 2002; Craig and Baihua, 1990; Craig and Busey, 2003). In normal-hearing adults, temporal order discrimination thresholds at the fingertips using brief vibromechanical stimuli are typically on the order of 20-40 ms. Eberhardt et al. (1994) used more sustained stimuli to examine temporal-onset order thresholds between vibratory and slow-displacement stimuli presented to the same finger, and found thresholds in the same range (25-45 ms).

When the paired stimuli delivered to the skin are brief and non-overlapping in time, judgments of temporal order may be based on several potential cues, including stimulus onsets and offsets, as well as total energy. By contrast, the sustained vibratory stimuli used in the present study allow us to examine onset- and offset-order discrimination capacities independently and to examine the contribution of total stimulus energy as a function of stimulus amplitude and duration.

Yuan and colleagues (Yuan, 2003; Yuan et al., 2004b, 2005a; Yuan and Reed, 2005) have examined both temporal onset-order and offset-order discrimination using sustained sinusoidal tactile stimuli, varying durations between 50-800 ms and amplitudes between 25-45 dB SL (relative to the average tactual detection threshold). Stimulus amplitudes and durations were chosen to span the ranges of these values previously observed among the amplitude envelopes of different frequency-filtered speech bands (Yuan et al., 2004a), which are used to modulate 50 Hz and 250 Hz vibrotactile carrier signals in the tactile speech coding scheme implemented by Yuan et al. (2005b) and Yuan and Reed (2005). Stimuli were presented in pairs—one was presented to the left thumb pad at a frequency of 50 Hz, and the other was presented to the left index finger pad at a frequency of 250 Hz—through the Tactuator, a multi-finger tactual display

(Tan & Rabinowitz, 1996). Different pulse rates were used for the two fingers in the interest of encoding location and vibratory frequency redundantly, with the aim of facilitating temporal order judgments (Taylor, 1978). Tactual temporal onset-order thresholds were consistent with those of previous studies, with thresholds averaging 34 ms across four normal-hearing young adults. The relative amplitudes of paired stimuli affected these thresholds far more substantially than the relative durations.

In contrast, offset-order thresholds were roughly four times as large as onset-order thresholds, averaging 142 ms across four normal-hearing young adults (Yuan et al., 2004b; Yuan and Reed, 2005). However, further studies of temporal offset-order discrimination in the tactual domain are lacking, presumably because temporal order has so routinely been examined using brief, non-overlapping stimuli.

Previous studies of tactual perception in deaf children have suggested some degree of enhanced performance, relative to normal-hearing controls, on a variety of tasks (e.g., Chakravarty, 1968; Schiff and Dytell, 1972), including two-point discrimination and line orientation discrimination. Cranney and Ashton (1982) reported enhanced tactile spatial discrimination in deaf subjects, relative to normal-hearing controls, in various age groups. Levanen and Hamdorf (2001) tested tactile sensitivity in congenitally deaf adults. They found that deaf subjects performed either comparably to or better than normal-hearing controls on two tasks involving tactual detection of changes in vibratory frequency in the range of 160-250 Hz.

Heming and Brown (2005) examined the perception of simultaneity of two tactile stimuli in subjects who had been profoundly deaf in both ears from before two years of age. Punctate mechanical stimuli were delivered, in pairs, to the pads of the index and middle fingers of either the left or right hand. Subjects were asked to indicate whether the two stimuli were "perceived simultaneously or non-simultaneously". Thresholds for perceived simultaneity on this task were significantly higher for deaf subjects (84 ± 25 ms) than for age-matched, normal-hearing controls (22 ± 15 ms). These results reflect pooled data from the left and right hands, which were similar to one another in both groups. Importantly, all subjects in this study were between 18-32 years of age. By contrast, in a previous study utilizing the same experimental technique, normal-hearing adults over 60 years of age had a mean threshold of 61 ms (Brown & Sainsbury, 2000), indicating that these simultaneity threshold values have strong dependencies that cannot be related simply to auditory experience. Furthermore, the perceptual judgment of "simultaneity" required of subjects in these studies is entirely subjective --- the stimuli are always presented with some amount of asynchrony --- and so differences in judgment criteria cannot be separated from the subjects' actual sensitivity. It should also be noted that normal-hearing subjects in Heming and Brown's study were

instructed in English, while profoundly deaf subjects were instructed in ASL. English and ASL are fundamentally different languages, and it is reasonable to consider that semantic factors might contribute to differences in decision criteria adopted by subjects in the two experimental groups. Given the subjective nature of Heming and Brown's simultaneity protocol, it is conceivable that their ten pre-lingually deaf subjects exhibited response tendencies that reflect social, cultural, and linguistic experiences that differ fundamentally from those of their normal-hearing counterparts. The question of tactual temporal resolution among congenitally deaf adults is examined using objective psychophysical methods in Chapter 3 of this document.

Chapter 2.

Tactile Speech Communication: Principles and Practice

2.1 Natural Methods of Tactual Speech Communication

2.1.1 The Tadoma Method of Tactual Speechreading

The feasibility of speech reception via the tactual sense is established by the Tadoma speechreading method, which has enabled some deafblind individuals to receive speech effectively by placing a hand on the speaker's face. Using tactile and kinesthetic cues, deafblind Tadoma-users can distinguish key articulatory features, including phonation, aspiration, and jaw/lip movements. Tadoma's practicability is limited by several factors, including requirements for extensive training and physical contact with a speaker. Nonetheless, Tadoma's efficacy in the hands of trained deafblind individuals has held up under scientific scrutiny.

The Tadoma user places his hand on the face of a speaker, with the fingers spread to detect cues of articulation. The hand is commonly oriented such that the thumb lightly touches the speaker's lips, though finger placement varies among users (Figure 2-1, top). Informative tactual cues include glottal vibration, positioning of the jaw, as well as movement and airflow at the lips (Reed et al, 1985). Positioning of the hand varies between Tadoma users, according to individual preferences, both of the user and the instructor. Some Tadoma users employ two hands. In some Tadoma variants, a finger placed at the speaker's nose can help to discern nasalized voicing, such as "m" and "n" sounds in English. Sometimes, an inverted hand position was used, with the thumb contacting the throat, as was the practice of Helen Keller (Figure 2-1, bottom). In the basic paradigm for training speech production, a child would hold her hand to the teacher's face during an utterance, and then she would hold her hand to her own face and try to reproduce this sensation using her own voice (Reed, 1995).

Beginning in the late 1970's, researchers at MIT took up a series of quantitative investigations into the efficacy of the Tadoma method, as practiced by experienced users. They sought primarily to assess how effectively various elements of speech were actually received and what benefits to language acquisition could be measured objectively in individuals for whom Tadoma had ostensibly worked well. The

participants were nine deafblind individuals, ranging in age from 31 to 67, all of whom were trained in Tadoma as children and were still using Tadoma for speech reception at the time of the study. Seven of the nine had become deafblind as the result of meningitis, between 18 months and 7 years of age. The remaining two participants were congenitally deafblind, with sensory impairment progressing in severity over infancy and childhood (Reed et al., 1985; Chomsky, 1986; Tamir, 1989; Reed, 1995).

An initial set of tests examined the perception of vowel and consonant constructs through Tadoma. Participants sat facing a speaker, placing one hand on the speaker's face in whatever arrangement they preferred. In both consonant and vowel tests, the participant responded on a Braille response form, indicating which sound they had perceived. In each trial of the consonant test, the speaker produced one of 24 possible consonants, always followed by the vowel /a/. All nine participants responded correctly in 52% to 69% of trials. In each trial of the vowel test, the speaker produced one of 15 possible vowel sounds, embedded between the consonants /h/ and /d/ (so if the vowel was /i/, then the utterance was pronounced "heed"). On this test, the results varied more widely between participants—their scores ranged between 18% and 60% correct.

Two further tests examined spoken word reception by the Tadoma users. In the "W-22 words" test, the speaker produces monosyllabic words in isolation—words such as "cars" and "tree"—taken from a list that is chosen to be phonetically-balanced. The participants responded by saying and then spelling each word, as they had perceived it. Their overall scores on this test ranged from 26% to 56% correct.

For the Speech Perception in Noise (SPIN) test, participants were required to identify the final word in a spoken sentence, again by repeating and then spelling their word choice (no additional "noise" was introduced, as suggested by the test's name). When the rest of the sentence provided no predictive information about the final word (e.g., "David does not discuss the hug."), the Tadoma users' accuracy fell in the range of 20% to 54% correct—quite similar to their accuracy with W-22 words presented in isolation. However, when the sentence did provide a predictive context for the final word (e.g., "All the flowers were in bloom."), the range widened to 24% to 86% correct. Four of the nine Tadoma users drastically improved their ability to recognize words when given the benefit of contextual information: three of these participants more than doubled their scores. The remaining five participants all showed very little or no benefit from sentence context, with accuracy again comparable to that achieved with words in isolation.

The nine Tadoma users were next tested on the CID sentences, which are designed to measure reception of "everyday" conversational speech at various speaking rates. The participants "listened" as their respective speakers produced these sentences one by one. After each sentence, participants would respond by repeating the sentence aloud. At a speaking rate of 2.5 syl/s (syllables per second), seven of the nine participants repeated the materials with between 65% and 85% accuracy, while the remaining two subjects both scored below 50% correct.

Three of the better performers from this test were examined further with similar materials spoken at faster rates. Two of these individuals demonstrated a sharp decline in speech reception for speaking rates above 3 syl/s, while the performance of the third did not show a similar decline until nearly 7 syl/s. Note that a normal speaking is about 4 to 5 syl/sec (Reed et al., 1985; Reed, 1995).

Overall, the researchers concluded from these tests that "the performance of Tadoma users may be compared to that of listeners in adverse speech-to-noise ratios or to lipreaders: that is, less-than-perfect segmental information appears to be combined with contextual cues to achieve reasonably good reception of conversational speech" (Reed, 1995, p. 43).

The three better performing participants, who were previously singled out for testing at faster speech rates, were further examined by Chomsky (1986), this time to assess their linguistic abilities. Chomsky reported that the Tadoma users demonstrated impressive syntactic abilities relative to the sighted deaf population. Furthermore, their vocabularies, as well as both oral and written language skills, proved comparable to those of typical normal-hearing adults.

Given the above findings, which demonstrate quantitatively just how effective Tadoma can be, one might expect the Tadoma method to play a major role in deafblind education today. Such is not the case. Schultz et al (1984) conducted a survey of institutions and programs for the hearing impaired and deafblind in the U.S. and Canada, to assess the prevalence of Tadoma usage in the education of the deaf and deafblind. They received responses from 376 out of the 750 institutions contact. The respondents included teachers, consultants, specialists, program administrators and university faculty. Out of 376 respondents, only 76 respondents (20%) had extensive knowledge of Tadoma. Another 102 respondents (27%) had heard of Tadoma, but had no further knowledge. The remaining 198 respondents (53%) indicated that they had never heard of the Tadoma method. Of the 76 respondents with Tadoma



Figure 2-1. Top panel: A conversation made possible by the Tadoma method. Two deafblind Tadoma users (left and right) communicate verbally with one another and a hearing, sighted individual (from Reed et al., 1985). Bottom panel: Helen Keller (left) and Anne Sullivan demonstrating the Tadoma method (using an alternative, inverted hand configuration) in the 1953 film *Helen Keller in Her Story*.

experience, only 36 respondents indicated that they were still using Tadoma with students in 1984, and most also used one or more additional methods of communication in conjunction with Tadoma. The most common of these other communication methods were auditory training/amplification and sign language, though fingerspelling, lipreading, and Braille were also represented (Schultz et al, 1984).

Perhaps the most significant factor identified with the decline of Tadoma is the increasing prevalence and severity of compounded physical and cognitive disabilities among deafblind individuals born during the past few decades. Gertrude Stenquist, who once taught Tadoma to deafblind children, observed that one reason for Tadoma's disappearance "is that so many children, born handicapped, survive these days who did not before, so that we are getting so many more children to whom we cannot teach speech. In fact, some of these little ones we have whom we are uncertain about, we may not be able to teach manually either". Infants born prematurely, though more likely to survive these days, are also far more likely to suffer sensorimotor and cognitive impairment. With this new deafblind population, the practicality of manual communication has finally overtaken the ideals of the oralist school.

Further reasons for Tadoma's exclusion from deafblind education are the extreme difficulty and time requirements of both teaching and learning speech through Tadoma. By contrast, a deafblind child can often acquire tactual sign language far more quickly and easily, and signing generally provides a more consistent and effective means of communication and instruction.

2.1.2 The Tactiling Method

Tactiling is a method of supplementing visual speechreading used by a relatively small number of hearing-impaired individuals. Tactual speech is received through a hand placed on the shoulders and throat of the speaker. It has the benefit of being less physically intrusive than Tadoma, as the deaf, sighted user does not touch the face of the speaker, but rather attends to visible speech cues. Although there is no standardized form of the Tactiling method, one variant of the strategy was formally evaluated by Plant and Spens (1986) as well as Ohngren et al. (1990). They evaluated the method of Tactiling as performed by a Swedish man (GS), who had become profoundly deaf at age eight due to meningitis, and had subsequently developed a tactually supplemented speechreading strategy through interactions with his parents. GS indicated that he found it most effective to hold his palm directly over a speaker's larynx, with fingers spread around the neck. However, due to social pressures, he had adapted the hand position such that his fingers rest on the speaker's shoulder while his thumb contacts the speaker's neck. Combined with visual speech-reading, this strategy quickly enabled GS to resume normal studies, which he completed through the university level. GS demonstrated (Swedish) speech-tracking scores (according to the tracking procedure described by DeFilippo and Scott, 1978) of over 60 words per minute (wpm) when Tactiling, compared with 40 wpm with visual speechreading alone. By comparison, tracking rates for normal-hearing Swedes average 83 wpm (Ohngren et al., 1990). Quite impressively, GS

demonstrated 99.3% accuracy in discriminating consonant voicing in /aCa/ context using the Tactiling method, compared with 61.1% accuracy via lipreading alone. He showed 91.7% accuracy in discriminating consonant manner using Tactiling, compared with 70.1% accuracy via lipreading alone. Within the category of articulatory manner, GS identified plosives with 88% accuracy via Tactiling, compared with 48.6% accuracy via lipreading alone (Plant and Spens, 1986). These results confirm the benefits of Tactiling for a highly experienced.

Ohngren and colleagues (1990) trained eight profoundly deaf adults to perform Tactiling over the course of five 2.5-hour training sessions, held once per week. Subjects were permitted to adjust their hand positions on the speaker's throat to a subjectively optimal position, and all of them chose to place the hand over the speaker's throat with the palm over the larynx. They found that Tactiling provides an immediate benefit to speechreading of sentences, improving the proportion of correctly recognized words by about 16%. No further benefit speechreading of sentences was observed following training. However, it should be noted that training was sparsely scheduled and involved only one speaker. Moreover, each of the five training sessions focused on discrimination among consonants in one of five visemic sets (e.g., /b/-/p/-/m/; /r/-/d/-/t/-/n/), with target consonants embedded in words and sentences, and with sessions spaced by one week. It is perhaps not surprising that a cumulative training benefit was not observed one week after the training of the final consonant set. Nonetheless, the immediate (pre-training) benefit of the Tactiling method for speech reception by profoundly deaf adults is substantial, and there remains the possibility that performance would improve with more adequate training.

2.2 Artificial Tactual Speech Communication

The concept of artificial tactual speech communication entered the scientific literature in the 1920's, with the work of R.H. Gault and colleagues. Gault initially experimented with the delivery of raw acoustic input directly to the skin of the hand through a tube, as well as the presentation of amplified speech through an earphone driver (Gault, 1924, 1926). With their Teletactor device, Gault and Crane (1928) filtered a speech signal into five frequency bands, and then presented the outputs to the fingers of deaf subjects via headphone drivers. However, since the skin's sensitivity to vibrations falls off quickly between 500-1000 Hz, much high frequency information was lost.

Beginning around 1948, a succession of researchers at MIT Research Laboratory of Electronics undertook a similar approach, applying a vocoder-type strategy to accommodate the tactual sense. Their Sensory Replacement system ("Felix") divided speech into five analog electronic channels through sharp-cuttoff

band-pass filters with bandwidths 0-400 Hz, 400-800 Hz, 800-1400 Hz, 1400-2400 Hz, and 2400-15000 Hz. The envelopes of the five bands were used to modulate the amplitudes of continuous 300 Hz sinusoidal signals, which served to drive five mechanical vibrators, upon which the five fingers of one hand could rest. Their initial report concerning speech reception indicates that "with about twelve sessions of training of ten minutes each, the average subject has been able to learn twelve words with 91 per cent accuracy". They next developed five- and seven-channel amplifiers with independently tunable filter band limits. By late 1949, they appear to have set aside the mechanical vibrators due to the requirement for bulky equipment, and instead turned to electrocutaneous stimulation, first at the fingertips, then on the forearm. Speech reception testing with the updated Felix system was initiated, but the project was dropped before publication (RLE Quarterly Progress Reports, 1949-50).

Many subsequent researchers and commercial developers of tactile aids have embraced this vocoder-type approach of dividing speech into distinct frequency bands, the envelopes of which are used to modulate vibration amplitudes on separate vibrotactile channels. Pickett and Pickett (1963) used a spatial array of ten vibrators, one applied to each finger of a subjects hands. Each of the ten channels carried a frequency band in a given region of the spectrum, with the total frequency range covered being 210-7700 Hz. The energy envelope of each band modulated the amplitude of a 300 Hz vibration. They tested two subjects on the pairwise discrimination of consonants and vowels. The average discrimination score across 14 vowel pairs was 83% correct, and the average score across 19 consonant pairs was 80%. Discrimination of vowels was found to be related to their distance in the F1-F2 vowel space. For consonants, voicing and nasality were detected on the basis of attack/decay characteristics, and articulatory place was also received.

Brooks et al. (1986) evaluated the Queen's University tactile vocoder, a 16-vibrator array presenting 100 Hz sinusoidal vibrations, with the signal at each vibrator being amplitude modulated by the energy envelope of a different frequency band between 160-8000 Hz. One subject trained to identify 250 words through the device alone (without lipreading), and was then tested with three different 1000-word open sets under the tactile, lipreading, and tactile+lipreading sensory conditions. Average word identification scores were 8.8% for tactile alone, 39.4% for lipreading alone, and 68.7% for lipreading+tactile. When the subject's responses were analyzed on the basis of accurate phoneme (rather than word) identification, the scores were 36.7% for tactile alone, 64.9% for lipreading alone, and 85.5% for lipreading+tactile. Moreover, the subject's word choices matched the actual stimulus word in syllabic stress pattern 88% of the time in the tactile alone condition, suggesting effective reception of certain prosodic information.

More elaborate displays have included two-dimensional vibrator arrays, in which adjacent vibrators along one dimension were typically driven by signals from adjacent frequency bands, while vibrators along the other dimension might reflect signal amplitude or provide a readout of recent activity within a given frequency band. For example, Clements et al. (1982) used an Optacon display, consisting of a 24 x 6 array of pins (approximately 2.8 mm x 1.2 mm), the heights of which are varied by individual cantilevered piezoelectric bimorphs. They tested two display strategies, both of which assigned frequency band-passed channels to the 24 rows of the array, but differed in usage of the six columns. A frequency-amplitude (FA) display encoded amplitude along the six columns, with the number of active pins in a column reflecting the corresponding channel amplitude. A time-swept (TS) display presented a binary-amplitude spectral representation along the length of a row, with each row representing successive 16-ms time frames. They tested the ability of two subjects' to discriminate pairs of vowels with each of the two displays, and found that both displays supported roughly 90% accuracy with synthetic vowels and roughly 80% accuracy with natural vowels in /b/-V-/t/ context.

Ifukube and Yoshimoto (1974) used a 16-channel filter with center frequencies ranging from 250-4000 Hz and bandwidths increasing with frequency. The display was a rectangular 16-by-3 arrangement of piezo-actuated reeds, which was applied to the fingertips. The energy envelope of each band modulated the amplitude of 200 Hz square-wave vibrations. They tested two encoding schemes, one of which simply presented the same signal on all three contacts in a column, and the other of which used a time-sweeping approach, with each of the three contacts reflecting a given input at a different latency. With the non-swept display, four normal-hearing subjects quickly learned to identify five vowels with 91% accuracy and five consonants with 66% accuracy on average. With half an hour of training, profoundly deaf children were able to identify the same five vowels with 85% accuracy on average. The deaf subjects were also tested with three sets of five consonants each. They averaged 35% accuracy in identifying the five plosives, 50% accuracy for the five nasal and glide consonants, and 65% accuracy for a set of five Japanese consonants chosen specifically because they are difficult to lipread. With the time-swept display, recognition of five CV syllables was found to improve by about 10% relative to performance with the non-swept display.

The MESA (multipoint electrotactile speech aid) device consisted of an 36-by-8 array of electrocutaneous contacts arranged along a belt. The 36 columns were used to code frequency of a speech signal, with band-pass filters having center frequencies ranging from 85 Hz to 10.5 kHz (and bandwidths increasing with center frequency). The eight contacts in each column provided an intensity code, with each contact assigned a 5 dB range of band signal amplitude, such that all eight electrodes covered a 40 dB range.

Sparks et al. (1978) trained subjects to identify consonants and vowels from among sets of phonemes produced by live talker. After 15 hours of training, one subject was able to identify vowels in an 8-vowel set with 95% accuracy in a single-talker scenario, and 76% accuracy when three talkers (two unfamiliar) were used. Confusions were mainly between the vowels /ɛ/, /æ/, /a/, and /ʌ/. After about 10 hours of training, two subjects identified plosive and nasal consonants with 50% accuracy on average. One subject also trained for eight hours on a set of nine fricatives, achieving 70% correct performance.

The limited efficacy of an assortment of speech transduction and tactual display strategies has defied researchers' basic intuitions. For example, whereas acoustic presentation of voice fundamental frequency (F0) substantially benefits accuracy and speed of lipreading in normal hearing listeners (Rosen et al., 1981; Boothroyd et al., 1988), numerous attempts to present F0 as a vibrotactile signal, using a variety of transduction strategies, have proven far less helpful (e.g., Eberhardt et al., 1990; Bernstein et al., 1989; but also see Boothroyd and Hnath-Chisolm, 1988). The simplest one- and two-channel devices have remained among the most effective of tactual aids in a running speech context, and no artificial device has yet matched the efficacy of the Tadoma speechreading method, as practiced by experienced deafblind individuals. The major practical successes in the development of artificial tactile speech systems have been devices that serve as supplements to lipreading.

2.3 Tactual Speech as a Supplement to Lipreading

Basic lipreading abilities vary substantially among hearing-impaired and normal-hearing individuals alike, both before and after periods of visual speech training. Visible facial cues convey little information about the laryngeal vibrations underlying voicing, which is a major disadvantage to a hearing-impaired individual when relying on lipreading to communicate (Heider and Heider, 1940; Walden et al., 1977; MacLeod and Summerfield, 1987; Dodd et al., 1989). While very few deaf individuals are able to understand speech fluently through lipreading alone, many can understand speech through lipreading supplemented by acoustic signals that convey aspects of speech not visible on the face (e.g., Rosen et al., 1981; Breeuwer and Plomp, 1984; Grant et al., 1991). For example, when they had normal-hearing listeners lipread short sentences, Breeuwer and Plomp (1984) found that the mean number of correctly perceived syllables increased by more than 40% when lipreading was supplemented with an acoustic tone, amplitude modulated by the energy envelope of a one octave band of speech centered at 500 Hz (which strongly reflects voicing amplitude).

Plant et al. (2000) compared the Tactiling method, in which a deaf individual places a hand on the speaker's shoulder/throat to supplement lipreading (described above), with lipreading supplemented by the amplified output of a contact microphone held against the speaker's neck presented through a hand-held bone conductor. The subject tested (GS) was highly experienced with the Tactiling method. From a set of 12 English consonants, GS correctly identified 50% via lipreading alone (LR), 82% via Tactiling (LRT), and 88% using the hand-held bone conductor as a supplement to lipreading (LRTac). The specific benefits of both the LRT and LRTac for discrimination of consonant voicing and articulatory manner were comparable to those observed by Plant and Spens (1986) for the Tactiling relative to lipreading alone (reviewed in section 2.1.2). From a set of five English vowels, GS correctly identified 53% via LR, 82% via LRT, and 71% via LRTac.

Plant and colleagues also tested GS in identification of monosyllabic words in English and Swedish (GS's native tongue). For English words, his percent correct scores were 18%, 34%, and 36% for the conditions LR, LRT, and LRTac, respectively. For Swedish words, his percent correct scores were 36%, 56%, and 42% for the conditions LR, LRT, and LRTac, respectively. Thus, the benefits of Tactiling over lipreading alone were substantial for both English and Swedish words, whereas the benefit derived from the artificial LRTac system was observed mainly with English words, for which his lipreading-alone skills were lacking.

Confusions among voiced and unvoiced consonants (e.g., /b/ and /p/) are particularly common in lipreading (Heider and Heider, 1940; Walden et al., 1977; Bratakos et al., 2001). In the interest of improving speech reception by deaf individuals, one strategy has been to reintroduce consonant voicing information to the lipreader as a vibrotactile temporal cue (Yuan et al., 2004a, 2005b). This strategy was demonstrated effective in enabling normal-hearing subjects to make consonant voicing discriminations when lipreading isolated syllables supplemented by the tactual cue (Yuan et al., 2005b), but the utility of tactual cue for deaf subjects was not examined. Utilization of the vibrotactile cue for consonant voicing requires a user to discriminate tactual temporal onset and offset asynchronies in the range of 50 to 200 ms. Of the few studies addressing tactual processing in deaf adults, the one most directly relevant to tactual temporal resolution (Heming & Brown, 2005) suggests that early deafness is associated with altered thresholds (or criteria) for reporting that two tactile stimuli have been presented simultaneously. Thus, a portion of the original research described in this document has examined tactual temporal resolution in congenitally deaf and normal hearing subjects using objective methods and stimuli that are chosen to approximate those used by Yuan et al. (2005b) to convey consonant voicing as a tactual cue.

Chapter 3.

Tactual Detection and Temporal Order Resolution in Congenitally Deaf Adults

3.1 OVERVIEW

During speech, visual facial cues to the voicing distinction are extremely weak, a major disadvantage to a hearing-impaired individual when relying on lipreading to communicate (Heider and Heider, 1940; Erber, 1974; Walden et al., 1977). Confusions between voiced and unvoiced consonants are common in lipreading, particularly when voicing is the only cue for discriminating between the consonants. In the interest of improving speech reception by deaf individuals, one strategy has been to reintroduce cues to consonant voicing to the lipreader as a vibrotactile temporal cue (Yuan et al., 2004a, 2005b). Such a temporal cue has been demonstrated effective in enabling normal-hearing subjects to make consonant voicing discriminations when lipreading isolated syllables supplemented by tactual cues (Yuan et al., 2005b), but the utility of such tactual cues for deaf subjects has not yet been examined. Utilization of the vibrotactile voicing cue requires a user to discriminate the temporal onset-and offset-ordering of tactual stimuli with asynchronies in the range of approximately 50 to 250 ms (Yuan et al., 2004a, 2004b). Of the few studies addressing tactual processing in deaf adults, the one most directly relevant to tactual temporal resolution (Heming & Brown, 2005) suggests that early deafness is associated with altered thresholds (or criteria) for reporting that two tactile stimuli have been presented simultaneously.

To guide further development of a strategy for tactile speech reception, tactual detection and temporal order discrimination by congenitally deaf and normal-hearing adults have been examined. Tactual detection thresholds for sinusoidal vibrations between 2 and 300 Hz were measured at the left thumb and index finger using an adaptive paradigm. Tactual temporal resolution was investigated using objective methods and stimuli that are chosen to approximate those used by Yuan et al. (2005b) to convey consonant voicing as a tactual cue. Specifically, temporal onset- and offset-order discrimination were tested using stimuli of 50-Hz at the thumb and 250-Hz at the index finger, delivered asynchronously and varied independently in amplitude and duration. Mean detection thresholds for the deaf and normal-hearing groups did not differ significantly at any frequency tested. Temporal onset-order discrimination

thresholds varied widely, particularly among congenitally deaf individuals, but no statistically significant difference was found between group means. Both experimental groups exhibited a broad range of discrimination thresholds for temporal offset-order, and mean thresholds did not differ significantly. On the whole, tactual offset-order thresholds were substantially higher than onset-order thresholds. Differences in the relative levels of paired stimuli systematically affected sensitivity to both onset- and offset-order in most subjects. Differences in the relative durations of paired stimuli had little effect on onset-order discrimination, but had a robust effect on offset-order discrimination thresholds, which was consistent across all subjects.

3.2 METHODS

3.2.1 Subjects

Fourteen right-handed individuals, ranging in age from 18 to 58 years, served as subjects in this study. Nine of these (five females and four males, ages 18 to 56) had been profoundly deaf in both ears from birth (Table 3-1). Their primary native languages included American Sign Language (ASL), Signed Exact English (SEE), Pidgin Signed English (PSE), Spoken English, Cued English, and Total Communication; several subjects reported more than one. The other five subjects (one female and four males, ages 23 to 58) had normal hearing (defined as 30 dB HL or better in the frequency range of 250 to 4000 Hz), and all reported English as their primary native language. Audiometric testing was conducted on each subject prior to participation in this study, and audiometric thresholds are presented in Table 3-2. All subjects received compensation for their participation.

Normal-hearing subjects were instructed in the requirements of experimental tasks by means of verbal communication and demonstration. Deaf subjects were instructed through a combination of verbal (lipreading) and written communication, demonstration, and in one case, through an ASL interpreter. Effective communication was easily established with all subjects. At the start of each temporal discrimination experiment, prior to the collection of any data, all subjects demonstrated near-perfect performance in practice trials with the experimental variable set well above threshold, indicating that they understood the requirements of each task.

Table 3-1. Information on Subjects with Congenital Deafness. Under Communication Methods:

PSE = Pidgin Signed English; ASL = American Sign Language; SEE = Signed Exact English.

<i>Subject</i>	<i>Age</i>	<i>Sex</i>	<i>Communication Methods</i>		<i>Onset / Etiology</i>
			<i>Early</i>	<i>Current</i>	
CD1	18	F	Cued English (w/ some signing)	Cued speech	congenital / unknown
CD2	28	F	PSE, then ASL	ASL and English	congenital / unknown (also has mild CP)
CD3	33	M	SEE to pre-teen, then ASL	ASL	congenital / unknown
CD4	42	F	Oral	ASL	congenital / unknown
CD5	42	F	Oral to age 19, then ASL	ASL	congenital / Rubella
CD6	45	M	Total Communication, ASL	ASL	congenital / unknown
CD7	47	M	Oral	ASL	congenital (premature birth)
CD8	51	F	Oral to age 18, then Total Communication	ASL and English	congenital / unknown
CD9	56	M	ASL	ASL and English	congenital / hereditary

3.2.2 Apparatus

The tactual stimulating device, called the "Tactuator", is illustrated in Figure 4-2 (of the next chapter). It consists of three rods that lightly contact the thumb, index finger, and middle finger of the left hand, in an arrangement that allows for a natural hand configuration. Each rod is moved independently by a head-positioning motor, which is controlled by a DSP board through a servomechanism. The apparatus is described in detail in Tan and Rabinowitz (1996) and Yuan (2003). Briefly, it is designed to deliver stimuli with frequencies from dc to around 400 Hz, spanning most of the cumulative receptive range of the four major types of somatosensory afferent receptors, which together mediate kinesthetic, flutter, vibratory, and cutaneous tactual percepts (Bolanowski et al., 1988). Each rod has roughly a 26 mm range of motion, meaning that it can reliably deliver stimuli with amplitudes ranging from absolute threshold (the smallest displacement that can be detected) to about 50 dB above threshold, throughout its useful frequency range. The effect of loading due to light finger contact is minimal (Tan and Rabinowitz, 1996).

Table 3-2. Audiometric thresholds of congenitally deaf (CD) and normal-hearing (NH) subjects in dB HL.

Thresholds in dB HL													
Subject	Age	250 Hz		500 Hz		1000 Hz		2000 Hz		4000 Hz		8000 Hz	
		Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
CD1	18	90	85	85	80	100	100	>120	115	120	120	>110	>110
CD2	28	80	>90	95	100	100	105	95	100	75	85	90	90
CD3	33	>85	85	>90	>90	>100	>100	>100	>100	>95	>95	-	-
CD4	42	>105	100	110	105	110	115	120	115	>120	>120	>110	>110
CD5	42	85	100	105	105	105	110	95	115	95	>115	>100	>100
CD6	45	75	85	75	80	85	90	85	85	65	70	105	110
CD7	47	55	50	100	75	95	90	90	85	85	90	80	85
CD8	51	85	>85	90	>85	>100	90	>100	>100	>95	>95	-	-
CD9	56	80	85	85	>90	>100	>100	>100	>100	>95	>95	-	-
NH1	23	0	0	5	5	0	0	10	5	10	5	0	5
NH2	30	15	10	5	5	0	0	5	10	5	15	15	15
NH3	35	5	10	0	0	0	10	5	10	15	30	20	20
NH4	45	5	5	15	5	5	5	15	0	15	10	40	30
NH5	58	5	5	5	5	5	5	5	5	20	10	45	30

3.2.3 Stimuli and Procedures

In all experiments, subjects sat facing a computer screen, which provided visual cues and feedback as appropriate. The subject's left hand rested on the tactual stimulating device, with the thumb and index finger gently touching the stimulators. The subject's right hand was situated by a computer keyboard, whereby responses were entered. During testing, both profoundly deaf and normal-hearing subjects wore foam earplugs (approximately 30 dB attenuating) and headphones delivering pink masking noise (roughly 80 dB SPL) in order to minimize the possibility of either air- or bone-conducted sounds from the device influencing performance at the highest stimulating levels.

Tactual detection threshold measurements

Sinusoidal vibratory stimuli were delivered to the distal glabrous surface of either the left thumb or the left index finger. Stimuli had 5-ms rise/fall times and were 500 ms in duration, sufficiently long such that detection threshold level should be independent of duration (Verrillo, 1965). The frequencies examined were 2, 5, 10, 25, 50, 100, 200, 250, and 300 Hz, which span most of the useful perceptual range of human tactual system for sinusoidal vibratory stimuli (Goff, 1967).

Detection thresholds were measured using a two-interval, two-alternative forced choice (2I-2AFC) adaptive procedure. At each finger, the order in which frequencies were tested was randomized. Subjects were asked to indicate in which of two visually-cued intervals a tactual stimulus was presented. Trial-by-trial correct-answer feedback was presented visually. Following an initial supra-threshold stimulus, subsequent stimulus presentations were governed by a "two-down, one-up" paradigm, which converges upon the stimulus level at which a subject would respond correctly 70.7% of the time (Levitt, 1971). An initial amplitude step-size of 4 dB was decreased to 1 dB following the second reversal in the direction of stimulus amplitude adjustment. Each run was terminated following the tenth reversal of direction, and the detection threshold was defined as the average stimulus level over the final six reversals. At each finger, and for each vibratory frequency, this protocol was repeated twice, and the results of the two runs were averaged.

Tactual temporal onset-order discrimination

To examine temporal onset-order discrimination, sinusoidal stimuli were delivered to the left thumb and forefinger during each trial of a one-interval, two-alternative forced choice (1I-2AFC) task, and the subject was asked to indicate which stimulus had the earlier onset. Subjects responded on a computer keyboard and received visual trial-by-trial correct-answer feedback. Subjects were instructed to press "T" if the onset of the stimulus delivered to the thumb preceded that of the stimulus delivered to the index finger, and to press "I" if the index finger stimulus onset arrived earlier. The thumb was stimulated exclusively at 50 Hz and the index finger at 250 Hz. Stimulus durations and amplitudes were varied in a manner that reflects duration and level variations across low and high-frequency bands of speech; this paradigm also ensured that neither offset asynchrony nor total stimulus energy could provide a reliable cue for onset-order. The duration of each stimulus in a trial pair was varied randomly among a fixed set of seven possible values: 50, 100, 200, 400, 500, 600 and 800 ms. Thus, there were 49 equally likely duration combinations for each stimulus pair. Signal levels at each site varied randomly among a set of five equally likely values. The 50 Hz stimuli to the thumb varied randomly among amplitudes of 26, 31, 36, 41 and 46 dB re 1 μ m peak displacement. The 250 Hz stimuli to the index finger varied randomly among amplitudes of 6, 11, 16, 21, and 26 dB re 1 μ m peak displacement. These amplitude values were previously selected by Yuan et al. (2005a) to span the range of 25-45 dB above the average detection thresholds of three normal-hearing adults (those thresholds being 1 dB re 1 μ m peak for 50 Hz at the left thumb and -19 dB re 1 μ m peak for 250 Hz at the left index finger). Beyond about 55 dB SL, vibratory stimuli become unpleasant or painful (Verrillo and Gescheider, 1992).

In the present study, average thresholds across all subjects are approximately 5 dB re 1 μ m peak for 50 Hz at the left thumb and -22 dB re 1 μ m peak for 250 Hz at the left index finger, and relative to these values, sensation levels ranged from about 21-41 dB for 50 Hz stimuli to the thumb and 28-48 dB for 250 Hz stimuli to the index finger. The stimulus amplitudes chosen by Yuan et al. (2005a) were retained for the sake of consistency, and their reference amplitudes for "sensation level" (dB SL) are retained throughout this text for purposes of clarity. Thus, the five stimulus amplitudes at both the left thumb and index finger are defined as having sensation levels of 25, 30, 35, 40, and 45 dB SL. Since both stimuli in a given trial were varied randomly among five possible amplitudes, there were a total of 25 possible amplitude combinations.

The amplitudes of the 50-ms and 100-ms duration stimuli were increased by 6 dB and 3 dB, respectively, relative to the other, longer stimuli. This adjustment is intended to compensate for the fact that temporal integration is observed for tactual stimuli of durations up to roughly 200 ms (Verrillo, 1965).

Two trials of the onset-order discrimination experimental paradigm are illustrated in Figure 3-1. We define stimulus onset asynchrony (SOA) as the difference in stimulus onset timings [$\text{Onset}_{\text{Thumb}} - \text{Onset}_{\text{Index}}$]. Thus, a negative SOA value ($\text{SOA} < 0$) in a given trial indicates that the thumb stimulus onset preceded that of the index finger, whereas a positive SOA value ($\text{SOA} > 0$) indicates that the index finger onset preceded that of the thumb. In Figure 3-1, "Trial 1" has a positive-valued SOA, since the thumb onset time is later than that of the index finger, whereas "Trial 2" has a negative-valued SOA --- the correct responses would be "I" in the first trial and "T" in the second trial. During any given run, the absolute value of the SOA was kept constant in each trial, and only the order of stimulus onsets (the sign of the SOA) was varied. Three SOA values were chosen for each subject during a pre-testing phase, so as to span the range for which the subject responded below 100% correct but above chance. This protocol reveals the pattern of temporal onset-order discrimination fall-off with decreasing $|\text{SOA}|$, and allows us to interpolate performance levels for intervening $|\text{SOA}|$ values. All subjects completed eight to ten 50-trial runs at each of the three selected values of $|\text{SOA}|$, which were interleaved randomly.

Tactual temporal offset-order discrimination

The temporal offset-order discrimination experiment was nearly identical to the onset order discrimination experiment, with the exception that the subject was now instructed to indicate which stimulus had a later offset (rather than an earlier onset). Subjects were instructed to press "T" if the offset of the stimulus delivered to the thumb was later than that of the stimulus delivered to the index finger, and

to press "I" if the index finger stimulus offset was later. All other aspects of the stimuli and the experimental paradigm were the same as those described above for the onset-order experiment.

Two trials of the offset-order discrimination experimental paradigm are illustrated in Figure 3-2. We define stimulus-offset asynchrony (SOFA) as the difference in stimulus offset timings [$\text{Offset}_{\text{Thumb}} - \text{Offset}_{\text{Index}}$]. Thus, a negative-SOFA value ($\text{SOFA} < 0$) for a given trial indicates that the index finger stimulus offset follows that of the thumb, whereas a positive SOFA value ($\text{SOFA} > 0$) indicates that the thumb offset follows that of the index finger. In Figure 3-2, "Trial 1" has a positive-valued SOFA, since the thumb offset time is later than that of the index finger, whereas "Trial 2" has a negative-valued SOFA --- the correct responses would be "T" in the first trial and "I" in the second trial. As in the onset-order experiment, the absolute value of the SOFA in the offset-order experiment was held constant for each trial of a given run, and only the sign of the SOFA was varied. Three SOFA values were chosen for each subject during a pre-testing phase, so as to span the range for which the subject responded below 100% correct but above chance. All subjects completed eight to ten 50-trial runs at each $|\text{SOFA}|$, which were interleaved randomly.

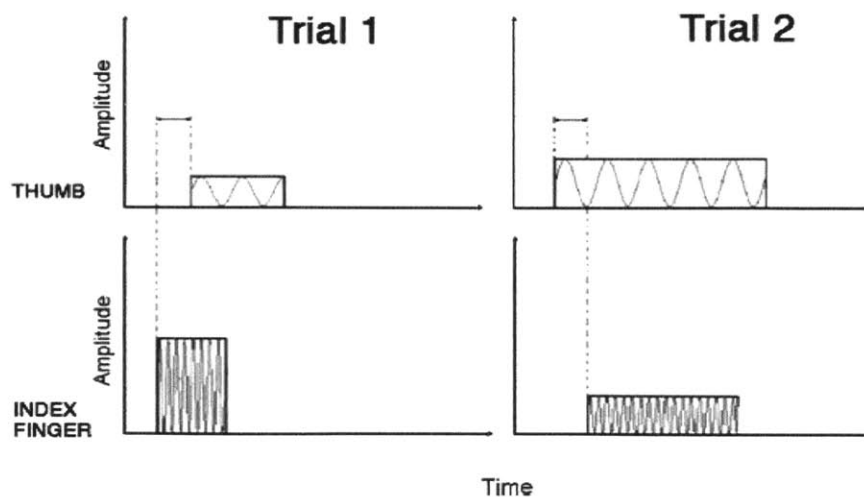


Figure 3-2. Two trials of the temporal onset-order experimental paradigm. Trial 1 has a positive-valued SOA (index onset precedes thumb) and Trial 2 has a negative-valued SOA (thumb onset precedes index). The upper traces represent 50-Hz vibration to the thumb, and the lower traces represent 250-Hz vibration to the index finger.

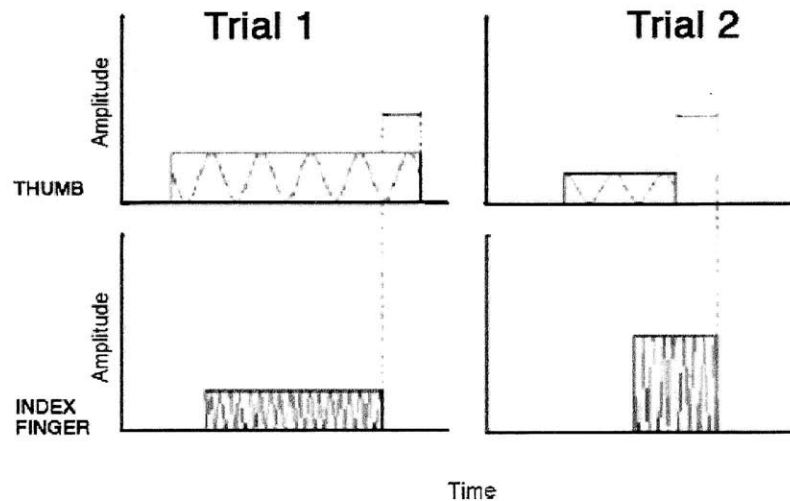


Figure 3-3. Two trials of the temporal offset-order experimental paradigm. Trial 1 has a positive-valued SOFA (thumb offset follows index) and Trial 2 has a negative-valued SOFA (index offset follows thumb). The upper traces represent 50-Hz vibration to the thumb, and the lower traces represent 250-Hz vibration to the index finger.

3.2.4 Data analysis

Tactual detection thresholds

The detection threshold in each experimental run was defined as the average stimulus level over the final six reversals of a two-down, one-up adaptive procedure. This protocol was repeated twice for each subject and averaged for each different combination of finger and vibratory frequency. Results were also averaged across subjects within each of the two groups (deaf and normal-hearing).

Tactual temporal onset-order and offset-order discrimination

Results from each experimental run of the temporal onset- and offset-order discrimination experiments for each subject were summarized in a 2x2 stimulus-response confusion matrix. The signal detection measures of sensitivity (d') and bias (β) were calculated (Green and Swets, 1966; Durlach, 1968) and averaged over all runs sharing a common $|\text{SOA}|$ or $|\text{SOFA}|$. Data from the onset- and offset-order experiments were summarized in plots of d' vs. $|\text{SOA}|$ and d' vs. $|\text{SOFA}|$, respectively, for each subject

separately. The discrimination threshold was then estimated as the interpolated value of $|\text{SOA}|$ or $|\text{SOFA}|$ corresponding to $d'=1$.

The effects of roving stimulus amplitudes and durations on temporal onset- and offset order discrimination were also examined. For the analysis of amplitude effects on temporal onset-order discrimination, data are grouped into five categories along the abscissa, according to the difference in sensation level between the earlier-onset and later-onset stimuli, described in Table 3-3. The sensation level of the earlier-onset stimulus is lower than that of the later-onset stimulus by 15-20 dB SL in trials of amplitude category 1, and by 5-10 dB SL in those of category 2. In trials of category 3, the two stimuli have approximately equal sensation levels. In trials of categories 4 and 5, the sensation level of the earlier-onset stimulus is larger than that of the later-onset stimulus by 5-10 dB SL and 15-20 dB SL, respectively. (Note that this system of categorization does not take into account the identities of the stimulated digits, but only the order of presentation.) For each subject, trials were sorted by amplitude category and $|\text{SOA}|$ value, and sensitivity (d') was calculated from 2x2 confusion matrices for each set of conditions separately.

For the analysis of stimulus duration effects on temporal onset-order discrimination, data are grouped into seven categories according to the difference between the durations of the earlier- and later-onset stimuli, as described in Table 3-4. In trials of categories 1-3, the earlier-onset stimulus is shorter in duration than the later-onset stimulus. In trials of category 4, the two stimuli have equal durations. In trials of categories 5-7, the earlier onset stimulus is longer in duration than the later-onset stimulus. (See Table 3-4 for specific duration difference criteria for each category.)

For the analysis of stimulus amplitude effects on offset-order discrimination, data are grouped into five categories along the abscissa of each graph, according to the difference in sensation level between the later-offset and earlier-offset stimuli, as described in Table 3-3. The sensation level of the later-offset stimulus is lower than that of the earlier-offset stimulus by 15-20 dB SL in trials of categories 1, and by 5-10 dB SL in those of category 2. In trials of category 3, the two stimuli have approximately equal sensation levels. In trials of categories 4 and 5, the sensation level of the later-offset stimulus is larger than that of the earlier-offset stimulus by 5-10 dB SL and 15-20 dB SL, respectively. For each subject, trials were sorted by amplitude category and $|\text{SOFA}|$ value, and sensitivity (d') was calculated from 2x2 confusion matrices for each set of conditions separately.

Table 3-3. Descriptions of categories used in analysis of stimulus amplitude effects on temporal onset-order and offset-order discrimination (Figure 3-6). The relationships indicated for paired stimuli in each amplitude category reflect frequency-specific sensation levels, rather than absolute amplitude differences.

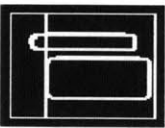
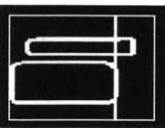
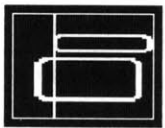
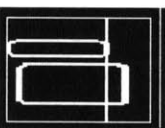
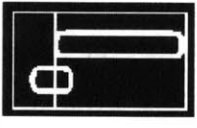
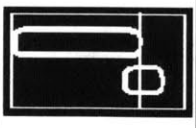
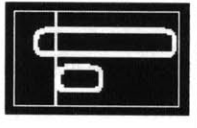
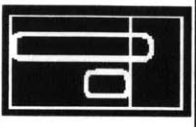
Amplitude Category	stimuli differ by	Onset-Order Experiment		Offset-Order Experiment	
1	15-20 dB SL	Earlier-onset stimulus has lower sensation level		Later-offset stimulus has lower sensation level	
2	5-10 dB SL				
3	0 dB SL	Stimuli have equal sensation levels		Stimuli have equal sensation levels	
4	5-10 dB SL	Earlier-onset stimulus has higher sensation level		Later-offset stimulus has higher sensation level	
5	15-20 dB SL				

Table 3-4. Descriptions of categories used in analysis of stimulus duration effects on temporal onset-order and offset-order discrimination (Figure 3-6).

Duration Category	stimuli differ by	Onset-Order Experiment		Offset-Order Experiment	
1	450-750 ms	Earlier-onset stimulus has shorter duration		Later-offset stimulus has shorter duration	
2	300-400 ms				
3	50-200 ms				
4	0 ms	Stimuli have equal durations		Stimuli have equal durations	
5	50-200 ms	Earlier-onset stimulus has longer duration		Later-offset stimulus has longer duration	
6	300-400 ms				
7	450-750 ms				

For the analysis of stimulus duration effects on offset-order discrimination, data are grouped into seven categories according to the difference between the durations of the later- and earlier-offset stimuli, as described in Table 3-4. In trials of categories 1-3, the later-offset stimulus is shorter in duration than the earlier-offset stimulus. In trials of category 4, the two stimuli have equal durations. In trials of categories 5-7, the later offset stimulus is longer in duration than the earlier-offset stimulus. (See Table 3-4 for specific duration-difference criteria for each category.)

3.3 RESULTS

3.3.1 Detection thresholds

Tactual detection threshold measurements for congenitally deaf (CD) and normal-hearing (NH) subjects are plotted as a function of stimulation frequency in the left- and right-hand panels of Figure 3-3, respectively. The upper panels of Figure 3-3 present data for each subject's left thumb, whereas the lower panels present data for the left index finger. Table 3-5 provides the mean thresholds and standard deviations of the two subject groups, calculated for each of the vibratory frequencies examined. Stimulus amplitude is reported in dB with respect to a 1 μm peak displacement.

Threshold measurements for all subjects exhibit the expected dependence on stimulation frequency. Thresholds at both fingers are lowest in the range of 200-300 Hz, increasing rapidly as frequency decreases below 200 Hz. A two-way analysis of variance (ANOVA) was performed within each frequency, testing for effects of group (deaf, hearing) and site of stimulation (thumb, index finger). No significant effects of either group or stimulation site were observed at any of the nine frequencies examined. The ANOVA for detection threshold measurements at 100 Hz gave the lowest p-values for the null hypotheses on the main effects for group, $F(1,24)=1.65$, $p=.21$, as well as for stimulation site, $F(1,24)=2.4$, $p=.13$.

Stimuli in the temporal order discrimination experiments were restricted to 50 Hz at the left thumb and 250 Hz at the left index finger. Mean (± 1 s.d.) detection thresholds for 50 Hz stimuli at the left thumb are 4.4 (± 7.5) dB re 1 μm for CD subjects and 5.9 (± 5.6) dB re 1 μm for NH subjects. Mean (± 1 s.d.) detection thresholds for 250 Hz stimuli at the left index finger are -22.9 (± 6.4) dB re 1 μm for CD subjects and -21.6 (± 4.8) dB re 1 μm for NH subjects.

3.3.2 Temporal onset-order discrimination

The results of the temporal onset-order discrimination experiment are presented in Figure 3-4, and onset-order thresholds (defined as the interpolated value of $|\text{SOA}|$ for which $d'=1$) are summarized in Table 3-6.

In the top panels of Figure 3-4, each data point indicates the mean and standard deviation of d' in runs with a given $|\text{SOA}|$, plotted for each subject individually. All subjects were tested at $|\text{SOA}|$ values of 50, 100, and 160 ms, with the exception of CD4, who was tested at $|\text{SOA}|$ values of 100, 160, and 250 ms. A straight line, crossing the origin, was fit to the data of each subject by linear regression. We define the discrimination threshold as the value of $|\text{SOA}|$ at which this line crosses $d'=1$.

Table 3-5. Mean tactual detection thresholds and standard deviations (in dB re 1 μm peak displacement) for congenitally deaf and normal-hearing subjects across all vibratory frequencies examined, for left thumb and left index finger.

Frequency	Left Thumb				Left Index Finger			
	Congenitally Deaf		Normal-Hearing		Congenitally Deaf		Normal-Hearing	
	Mean (dB re 1 μm)	s.d.	Mean (dB re 1 μm)	s.d.	Mean (dB re 1 μm)	s.d.	Mean (dB re 1 μm)	s.d.
2	42.0	3.9	43.2	2.6	42.7	3.6	43.3	3.6
5	37.5	4.6	37.4	4.0	37.9	2.8	37.1	5.8
10	32.0	3.4	31.8	1.9	33.2	2.6	32.1	5.6
25	24.9	7.7	26.5	3.2	24.3	3.7	23.6	5.4
50	4.4	7.5	5.9	5.6	3.2	5.8	6.3	6.8
100	-11.4	7.0	-6.9	4.6	-14.3	7.8	-12.7	6.9
200	-22.8	5.0	-21.2	2.5	-21.9	8.8	-22.0	3.1
250	-24.1	6.5	-27.1	7.3	-22.9	6.4	-21.6	4.8
300	-25.2	5.7	-23.9	5.7	-21.8	5.7	-21.5	5.8

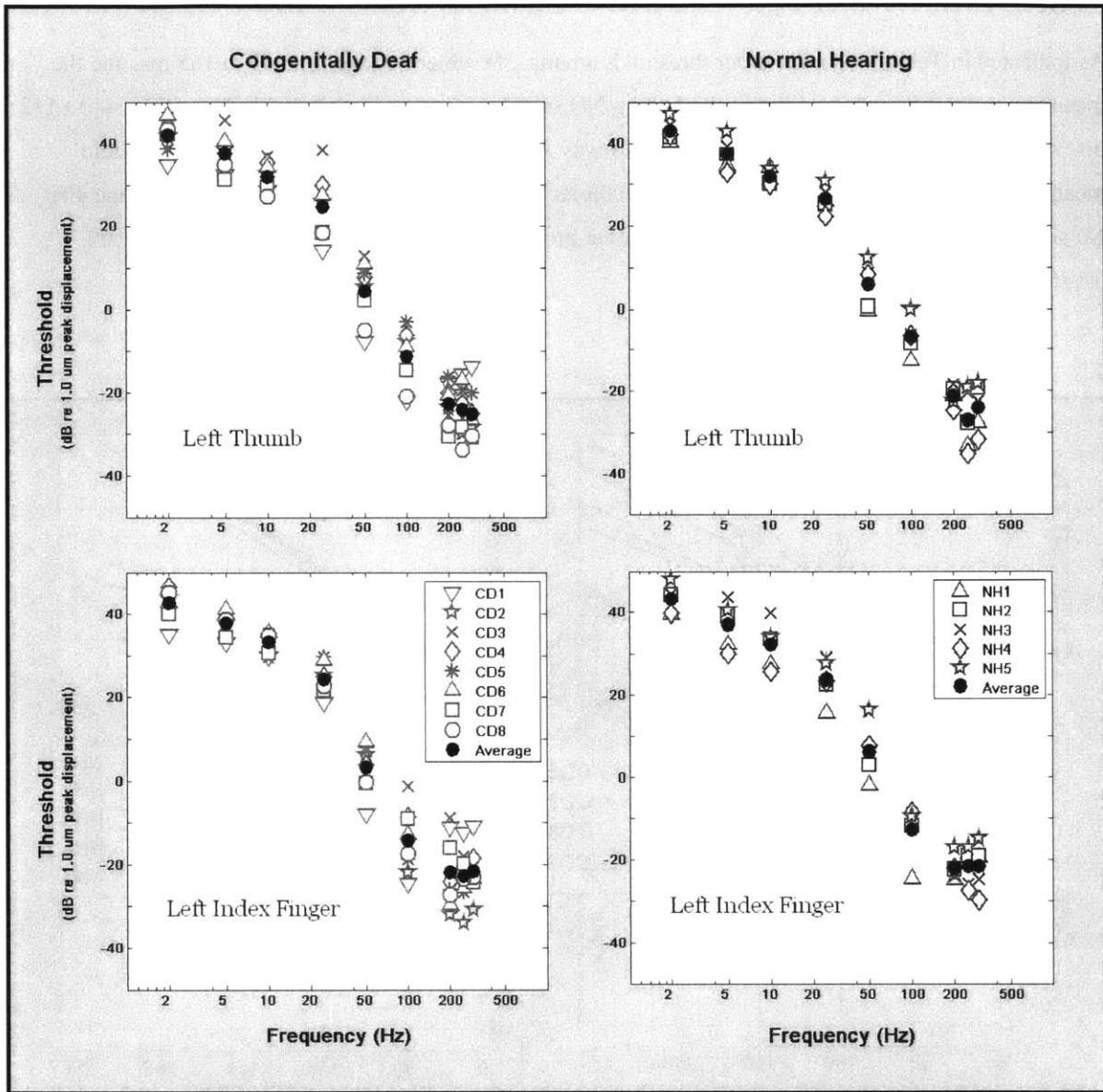


Figure 3-4. Tactual detection thresholds (in dB re 1.0 μm peak displacement) are plotted as a function of stimulation frequency. Measurements for congenitally deaf subjects are presented in the two left-hand panels and those for normal-hearing subjects are presented in the two right-hand panels. Data in the upper panels correspond to the left thumb of each subject, while those in the lower panels correspond to the left index finger.

As indicated in Table 3-6, onset-order thresholds among CD subjects range from 47 to 165 ms, and the mean threshold is 83 ms (± 41 ms s.d.). Among NH subjects, onset order thresholds range from 41 to 112 ms, with a mean of 58 ms (± 31 ms s.d.). A two-way ANOVA was performed to compare threshold means with respect to hearing status and age (subjects were categorized as either over or under age 40). No significant effects were demonstrated for either group, $F(1,10)=1.06$, $p=.33$, or age, $F(1,10)=.09$, $p=.76$.

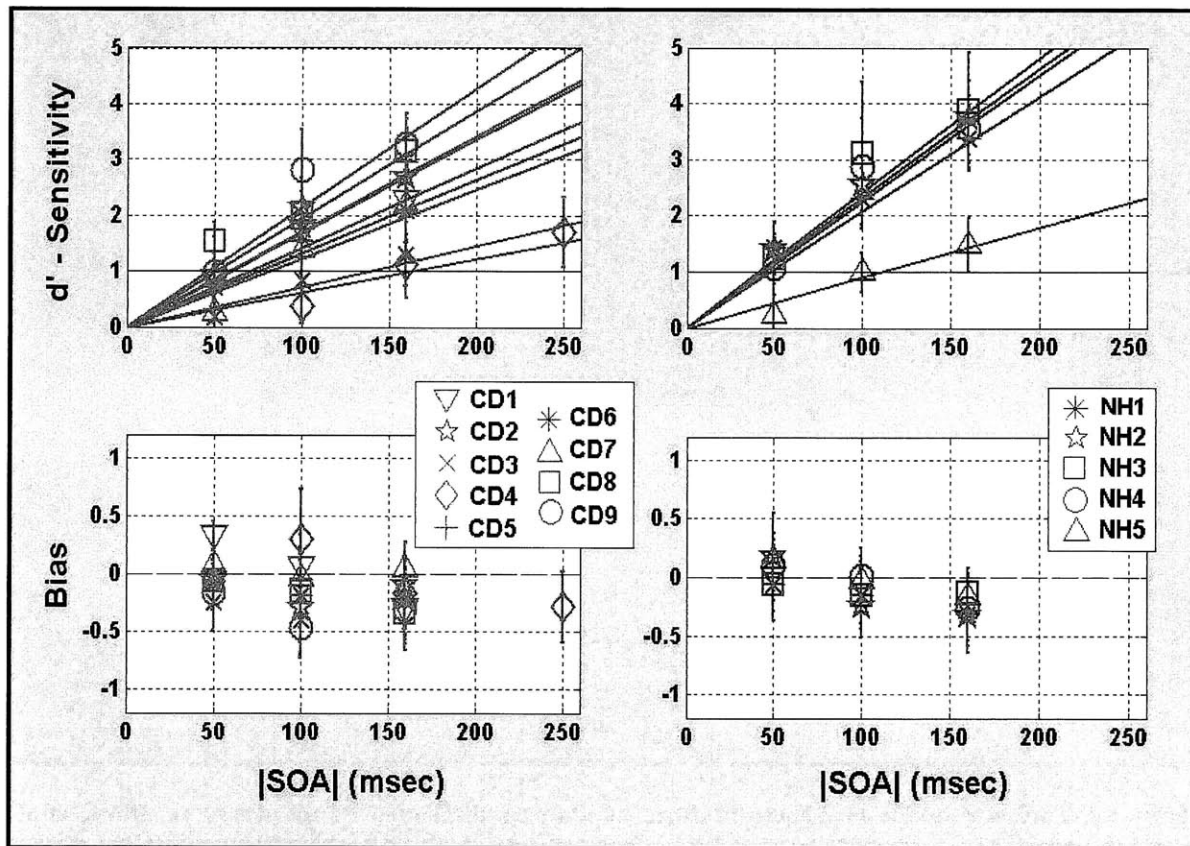


Figure 3-5. Mean values of d' (top) and bias (bottom) are plotted as a function of $|SOA|$ for congenitally deaf (left) and normal-hearing (right) subjects in the temporal onset-order discrimination experiment. Each subject was tested at three $|SOA|$ values. In the top panels, a straight line that crosses the origin has been fit to the data of each subject by linear regression; onset-order discrimination threshold is defined as the value of $|SOA|$ at which this line crosses $d'=1$. Error bars indicate standard deviation.

The corresponding bias (β) values are plotted in the bottom panels of Figure 3-4; each data point represents the mean bias over all runs of the same |SOA| for a given subject. Positive bias indicates a tendency to respond that the stimulus at the thumb had an earlier onset. For both CD and NH subjects, bias is generally negligible. In no case did $|\beta|$ exceed 0.5.

3.3.3 Temporal offset-order discrimination

The results of the temporal offset-order discrimination experiment are presented in Figure 3-5, and offset-order thresholds (defined as the interpolated |SOFA| at which $d'=1$) are recorded in the rightmost column of Table 3-6.

In the top panels of Figure 3-5, each data point indicates the mean and standard deviation of d' in runs with the corresponding |SOFA|, for a given subject. Most subjects were tested with |SOFA| values of either 50, 100, and 160 ms or 100, 160, and 250 ms; the only exception was subject CD4, who was tested with 160, 250, and 400 ms |SOFA| values. A straight line, crossing the origin, was fit to the data of each subject by linear regression, and the discrimination threshold is defined as the value of |SOFA| at which this line crosses $d'=1$.

Thresholds range from 75 ms to 276 ms among CD subjects, and the mean threshold is 139 ms (± 66 ms s.d.). Among NH subjects, thresholds range from 65 ms to 221 ms, with a mean of 121 ms (± 63 ms s.d.). A two-way ANOVA was performed to compare threshold means with respect to hearing status and age (subjects were categorized as either over or under age 40). No significant effects were demonstrated for either group, $F(1,9)=.12$, $p=.74$, or age, $F(1,9)=.19$, $p=.67$.

The corresponding β values are shown in the bottom panels of Figure 3-5; each data point indicates the mean bias over all runs of the same |SOFA| for a given subject. Positive bias indicates a tendency to respond that the stimulus at the thumb had a later offset. For CD and NH subjects, bias is minimal overall. In no case did $|\beta|$ exceed 0.5.

Table 3-6. Stimulus onset-order (|SOA|) and offset-order (|SOFA|) sensitivity thresholds (interpolated at $d' = 1$) for congenitally deaf and normal-hearing subjects.

Group	Subject	Age	Onset-Order Threshold (ms)	Offset-Order Threshold (ms)
Congenitally Deaf	CD1	18	70	132
	CD2	28	59	92
	CD3	33	138	183
	CD4	42	165	276
	CD5	42	59	-
	CD6	45	76	129
	CD7	47	82	83
	CD8	51	51	75
	CD9	56	47	142
	Mean ± s.d.		83 ± 41	139 ± 66
Normal-Hearing	NH1	23	48	145
	NH2	30	41	90
	NH3	35	44	83
	NH4	45	43	65
	NH5	58	112	221
	Mean ± s.d.		58 ± 31	121 ± 63

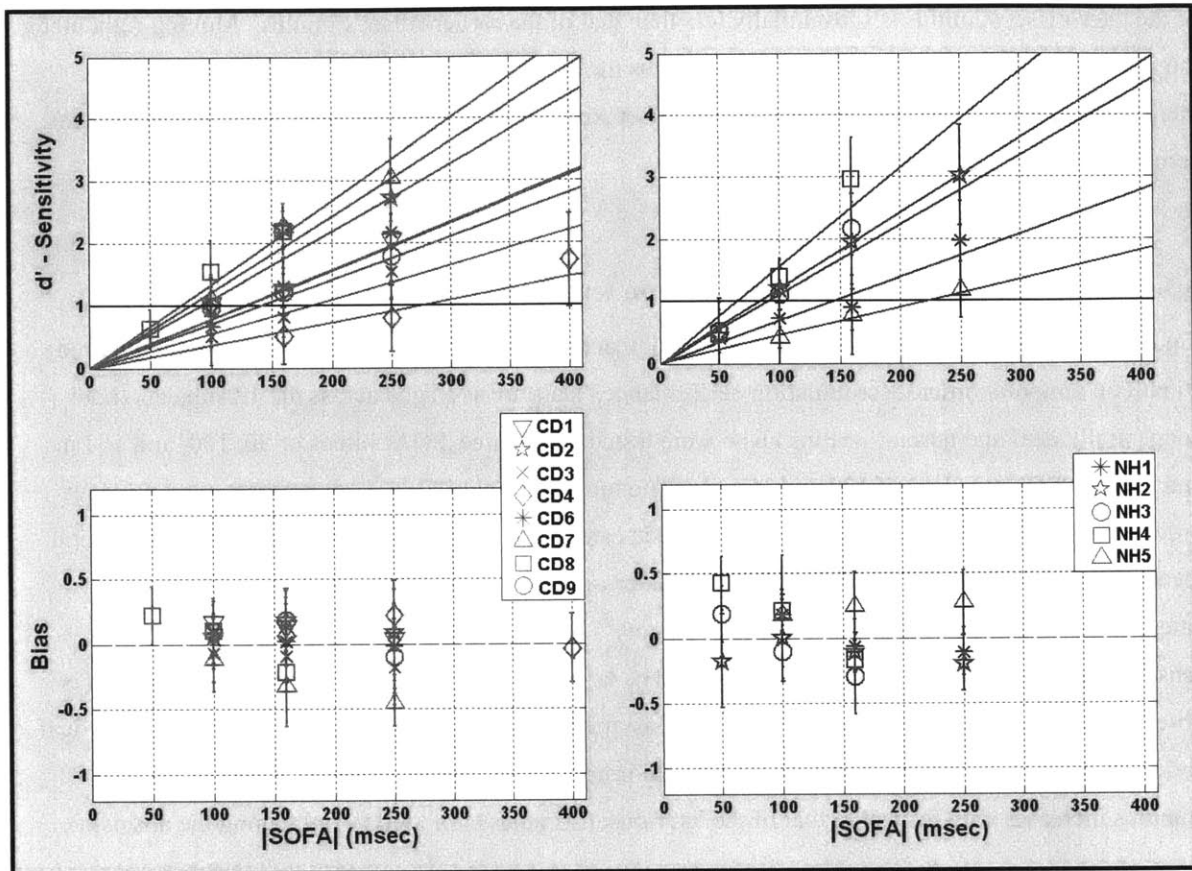


Figure 3-6. Mean values of d' (top) and bias (bottom) are plotted as a function of $|SOFA|$ for congenitally deaf (left) and normal-hearing (right) subjects in the temporal offset-order discrimination experiment. Each subject was tested at three $|SOFA|$ values, except NH2, who was tested at four. In the top panels, straight lines that cross the origin have been fit to the data of each subject by linear regression; offset-order discrimination threshold is defined as the value of $|SOFA|$ at which this line crosses $d'=1$. Error bars indicate standard deviation.

The bottom right panel of Figure 3-6 presents average offset-order thresholds plotted as a function of duration category. For the offset-order discrimination experiment, trials of duration category 1 are those in which the later-offset stimulus is shortest in duration relative to the earlier-offset stimulus --- with increasing duration category number, the duration of later-offset stimulus increases relative to that of the earlier-offset stimulus (as described in Table 3-4). In contrast to the findings of the onset-order discrimination experiment, relative stimulus duration has a very consistent and substantial effect on the discrimination of offset-order in normal-hearing and congenitally deaf subjects alike. As observed in the figure, subjects generally perform best in trials of low-number duration categories, for which the duration

of the later-offset stimulus is substantially less than that of the earlier-offset stimulus. Moving right along the abscissa, performance progressively worsens as the duration of the later-offset stimulus becomes increasingly long relative to that of the earlier-offset stimulus. This performance pattern was consistent across all subjects and |SOFA| values examined.

3.3.4 Effects of relative level and duration on temporal resolution

Figure 3-6 provides an overall impression of the impact of level and duration differences among paired stimuli on temporal order discrimination performance. Data are averaged across the 12 subjects (both congenitally deaf and normal-hearing) who were tested at the three |SOA| values of 50, 100, and 160 ms and the two |SOFA| values of 100 and 160 ms. The top left panel of Figure 3-6 presents average onset-order thresholds plotted as a function of "amplitude category". As described in Table 3-3, the sensation level of the earlier-onset stimulus relative to the later-onset stimulus is lowest in trials of amplitude category 1 and highest in trials of amplitude category 5 (whereas the paired stimuli have comparable sensation levels in amplitude category 3 trials). The largest difference between amplitude categories is observed for the 50-ms |SOA|, the smallest onset asynchrony tested. As a general trend among both deaf and normal-hearing subjects, performance tends to improve as the relative level of the earlier-onset stimulus increases with respect to that of the later-onset stimulus (from left to right along the abscissa).

Average offset-order thresholds for stimulus offset-asynchronies of 50 and 100 ms are plotted as a function of amplitude category in the bottom left panel of Figure 3-6. The sensation level of the later-offset stimulus relative to the earlier-offset stimulus is lowest in trials of amplitude category 1 and highest in trials of amplitude category 5 (as described in Table 3-3). Stimulus offset-order discrimination performance tended to improve as the relative level of the later-offset stimulus increased with respect to that of the earlier-offset stimulus (from left to right along the abscissa). This trend was observed consistently across all subjects, at all offset-order asynchrony values tested, although performance varied most substantially across amplitude categories in trials where the |SOFA| value was lowest, and performance was thus poorest overall.

In the top right panel of Figure 3-6, average onset-order thresholds are plotted as a function of "duration category". As described in Table 3-4, the duration of the earlier-onset stimulus relative to the later-onset stimulus is smallest in trials of duration category 1 and largest in trials of duration category 7 (whereas the paired stimuli have comparable durations in category 4 trials). Duration differences among paired stimuli are seen to have little, if any, effect on performance at any the three |SOA| values.

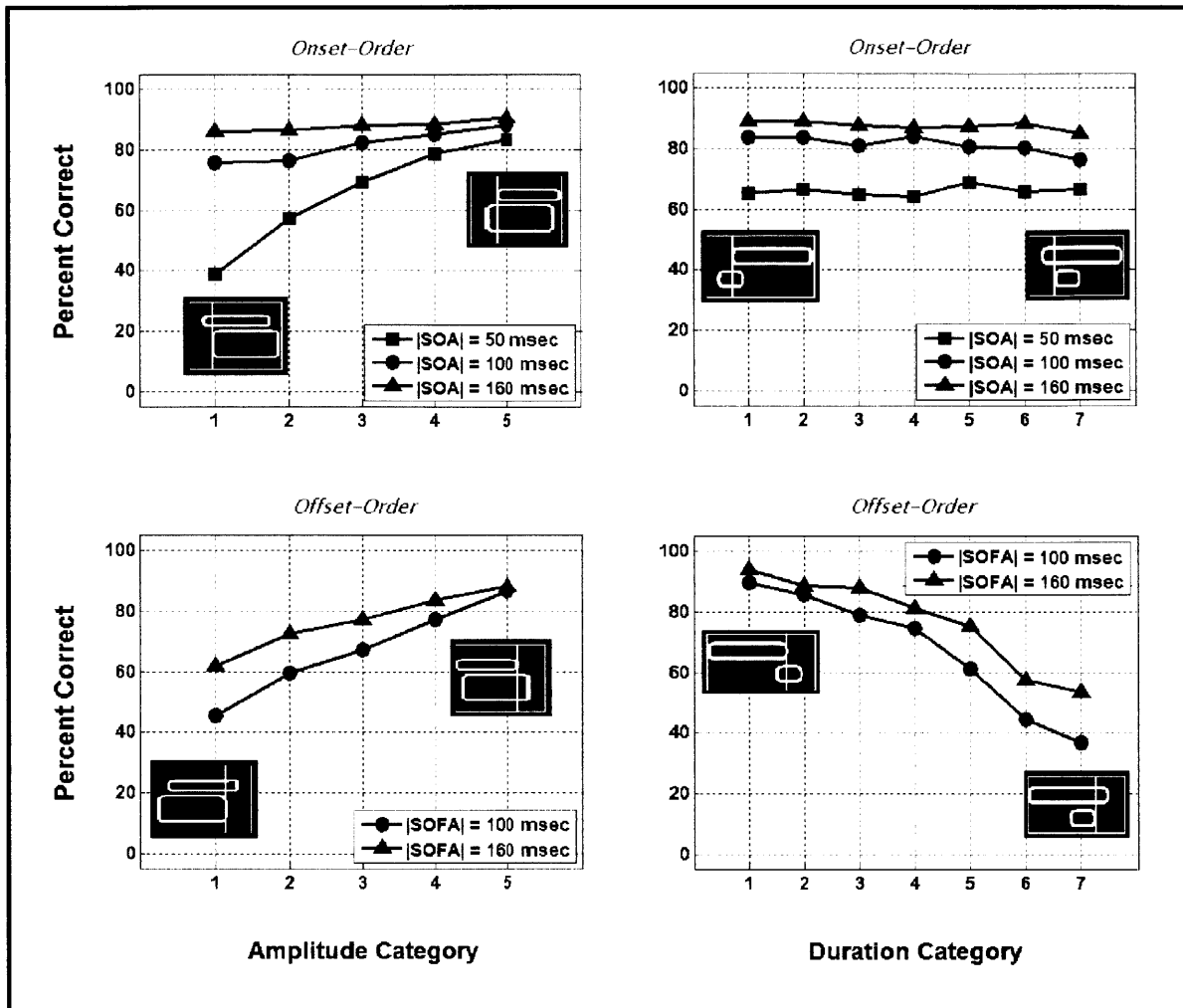


Figure 3-7. Percent correct performance plotted as a function of amplitude category (left panels) and duration category (right panels) in the tactual temporal onset-order and offset-order experiments (top and bottom panels, respectively). Data are averaged across the 12 deaf and normal-hearing subjects who were tested with stimulus onset asynchrony (SOA) values of 50, 100, and 160 ms and stimulus offset asynchrony (SOFA) values of 100 and 160 ms. In each panel, illustrative icons depict either the relative level or duration relationships characteristic of paired stimuli in the lowest and highest numbered categories, in conjunction with indicators of their relative temporal onset or offset ordering. The precise relationships characteristic of each amplitude and duration category are described in Tables 3-3 and 3-4, respectively.

3.4 DISCUSSION

3.4.1 Tactual detection

We examined tactual detection thresholds at the left thumb and index finger using sinusoidal stimuli at nine frequencies between 2-300 Hz. We found no significant differences, at any frequency, between the mean detection threshold of CD and NH subjects. Overall, our measurements are consistent with those reported in previous tactual sensitivity studies that used comparable experimental conditions (e.g., Yuan, 2003; Tan, 1996; Rabinowitz et al., 1987; Lamore et al., 1986).

Subjects in this study were free at all times to adjust the amount of pressure applied to the tactual stimulators by each finger. The contact area between each finger and the corresponding stimulator bar is approximately 0.5 - 1 cm², depending on finger size and positioning. Subjects were instructed regarding hand placement and told to touch the bars lightly, but any effects on threshold measurements resulting from the amount of skin indentation may have varied among subjects and between experimental sessions. Researchers who have examined tactile detection thresholds on the hand using a static indentation at the contact surface (e.g., Bolanowski et al., 1988; Gescheider et al., 2002) have observed a 10-20 dB reduction in thresholds at low frequencies (in the range of 1 to 25 Hz), whereas thresholds in the range of 50-300 Hz were elevated by about 10 dB, relative to studies that did not impose a static indentation. Our detection threshold measurements at both low and high frequencies, however, are more consistent with previous studies in which static indentation was not employed (e.g., Lamore et al., 1986; Rabinowitz et al., 1987).

3.4.2 Tactual temporal order discrimination

Tactual onset-order discrimination thresholds, listed in Table 3-6, varied widely among both CD and NH adults in this study. Onset-order discrimination thresholds in the CD group ranged from 47 to 165 ms (a factor of 3.5), and those in the NH group ranged from 41 to 112 ms (a factor of 2.7). The highest tactual temporal onset-order discrimination thresholds reported here exceed those typically observed among normal-hearing subjects (Sherrick, 1970; Eberhardt et al., 1994; Yuan et al., 2005a). In particular, Yuan et al. (2005a), who measured onset-order thresholds in four normal-hearing adults (aged 21- 32), used the same experimental paradigm as was used in the present study, and reported |SOA| thresholds ranging from 18 to 42 ms. Analysis of variance revealed no statistically significant difference between mean tactual onset-order thresholds of subjects grouped either by age or hearing status.

Several previous studies have demonstrated altered tactual temporal processing ability with aging (e.g., Van Doren et al., 1990; Gescheider et al., 1992). Elevated thresholds for judgments of simultaneity were previously observed in adults over 60 years of age (Brown & Sainsbury, 2000), suggesting a correlation between age and temporal resolution. The current results, however, do not indicate any trend of age-related elevations in onset- or offset-order thresholds, at least over the age range of 18-58 years examined here. Correlation analysis gives coefficients $r=0.133$ ($p=.65$) for onset-order threshold vs. age and $r=0.137$ ($p=.66$) for offset-order threshold vs. age.

Tactual offset-order discrimination thresholds for both CD and NH adults, listed in Table 3-6, span an even wider range of values than onset-order thresholds. Offset-order discrimination thresholds in the CD group ranged from 75 to 276 ms (a factor of 3.7), and those in the NH group ranged from 65 to 221 ms (a factor of 3.4). These threshold ranges are consistent with previously observed offset-order thresholds in normal-hearing subjects (Yuan et al., 2004b; Yuan and Reed, 2005).

Figure 3-7 shows tactual offset-order thresholds plotted as a function of corresponding tactual onset-order thresholds for each subject who participated in both temporal order discrimination experiments. For a given subject, offset-order threshold is generally substantially larger than onset-order threshold (although note that the two thresholds are about equal for subject CD7). Overall, subjects' offset-order discrimination thresholds are highly correlated with their respective onset-order thresholds (correlation coefficient $r = 0.852$, $p<0.0005$). Subjects NH5, CD3 and CD4, who have the three highest |SOA| thresholds in this study, are also found to have the highest |SOFA| thresholds.

In the auditory system, temporal offset-order discrimination thresholds for pairs of tonal stimuli are typically as low as (or lower than) onset-order thresholds (Pastore, 1983). Substantially higher offset-order thresholds observed for paired vibrotactile stimuli in this study may reflect, at least in part, the characteristically broad spatiotemporal integration patterns of the Pacinian system (which is engaged by moderate-level 50-Hz as well as 250-Hz sinusoidal vibrations). Interactions among vibrotactile stimuli that activate Pacinian receptors at two distinct skin locations have most commonly been investigated as masking phenomena. For example, Sherrick (1964) reported an increase of almost 30 dB in the detection threshold for a 350-ms pulsed 150-Hz vibrotactile signal at the right index finger when presented simultaneously with a 350-ms pulsed 150-Hz vibrotactile masker of 30 dB SL at the right little finger. Verrillo et al. (1983) observed substantial masking of a 300-ms, 300-Hz test stimulus at the index

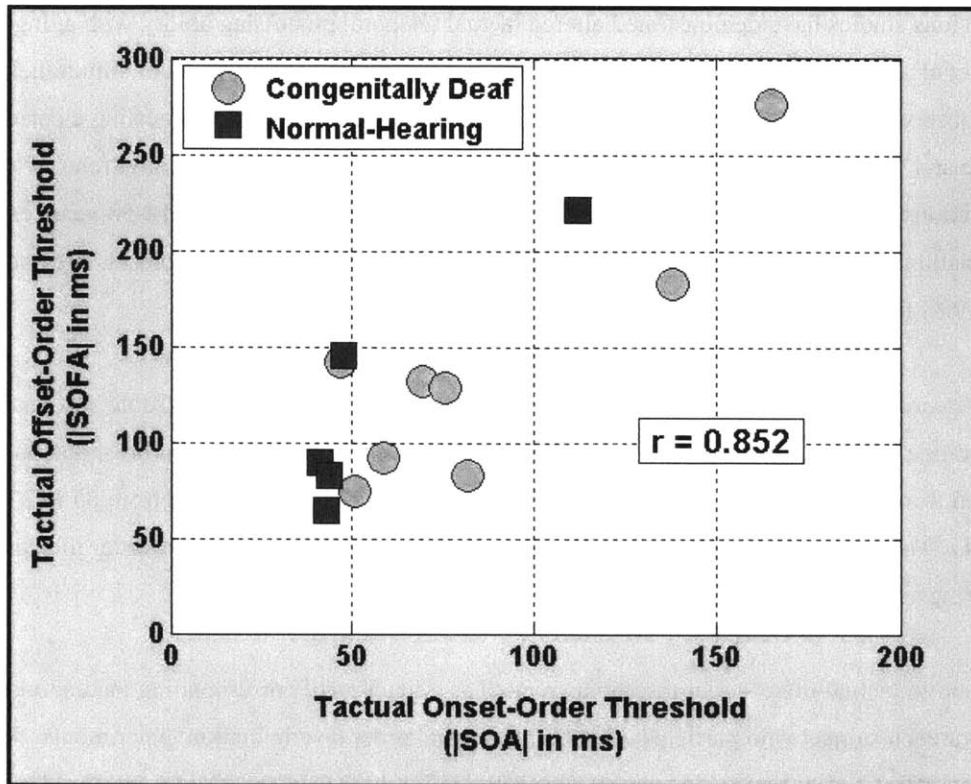


Figure 3-8. For each subject who participated in both temporal order discrimination experiments, tactual offset-order thresholds are plotted as a function of corresponding tactual onset-order thresholds. (Correlation coefficient $r = 0.852$, $p < 0.0005$)

fingertip when it was temporally centered within a 730-ms, 300-Hz masking stimulus at the ipsilateral thenar eminence. They further measured vibration at the index fingertip due to physical transmission of masking vibration from the thenar site, and thereby established that even their maximum intensity masker produced only subthreshold vibration at the fingertip surface (suggesting that the underlying process is likely more central). Given the marked ability of such stimuli to interfere with one another's detection, it seems reasonable that two partially-overlapping, Pacinian-range, vibrotactile stimuli delivered to adjacent digits might interact in such a manner as to obscure their relative temporal order. The fact that discrimination thresholds for offset-order tend to exceed those for onset-order suggests that the effects of vibrotactile forward masking extend over a longer time period than those of backward masking. Such a temporal asymmetry of vibrotactile forward and backward masking has previously been characterized as a "persistence" of tactual stimuli following their offset (Craig and Evans, 1987).

Heming and Brown (2005) previously examined thresholds for perceived simultaneity of two punctate stimuli delivered to separate fingers, among adults with early hearing loss as well as normal-hearing adults. The present study differs from theirs in several key respects. First, we used an objective experimental protocol in which subjects indicate the perceived ordering of stimuli, which enables us to apply signal detection theory to distinguish sensitivity from response bias. Second, we used pairs of sinusoidal stimuli, which were varied independently in amplitude and duration, thereby allowing us to assess the impact of these stimulus parameters on temporal resolution. Third, our use of sinusoidal (rather than punctate) stimuli has enabled us to examine separately the discrimination of relative onset and offset times.

Heming and Brown reported mean simultaneity thresholds of 84 ms (s.d.=25) and 22 ms (s.d.=15) for their deaf and normal-hearing subject groups, respectively. These results were interpreted as evidence of compromised tactile temporal processing in adults with early hearing loss. The results of the current study show a similar trend for larger mean thresholds for CD compared to NH subjects, but this trend does not reach significance. Tactual temporal onset order thresholds among individual CD subjects in the current study varied substantially, some being within the typical range for most normal-hearing subjects. Moreover, tactual temporal offset-order thresholds in the current study were comparable between the CD and NH groups.

The specific neural processing requirements of a temporal order discrimination task are likely somewhat different from those of an asynchrony detection task, which may, to some extent, account for disparities between our findings and those of Heming and Brown (2005). It has been shown, for example, that perceptual learning following multi-hour training on either an auditory temporal onset-order discrimination task or an auditory asynchrony detection task does not generalize from one to the other (Mossbridge et al., 2006). Moreover, the same study found no benefit to auditory temporal offset-order discrimination following training on either the onset-order or asynchrony task. It is thus likely that no single psychophysical metric can adequately support generalizations about sensory temporal processing in adults with early hearing loss.

3.4.3 Amplitude and duration effects

In the temporal onset- and offset-order discrimination experiments, the amplitudes and durations of paired stimuli were varied independently, spanning value ranges corresponding to those observed among the amplitude envelopes of different frequency-filtered bands of speech (Yuan et al., 2004a). Thus, the

experimental conditions employed in the present study roughly simulate the temporal stimulation patterns exhibited by vibrotactile carrier signals when they are modulated by filtered speech envelopes, as in the tactile speech coding scheme implemented by Yuan et al. (2005b) and Yuan and Reed (2005). We report here that the relative levels of two vibrotactile stimuli exert substantial influence over discrimination of both onset and offset asynchronies. Generally, as the sensation level of the stimulus at one site increases relative to the other, subjects are more likely to perceive that stimulus as having an earlier onset than the other stimulus in the SOA discrimination experiment, or as having a later offset in the SOFA discrimination experiment.

By contrast, differences in relative stimulus duration predominantly affect offset-order discrimination performance only. Specifically, when one stimulus in a pair is substantially shorter in duration than the other, subjects are more likely to perceive that stimulus as having a later offset. Consider that, in the SOA experiment, an increase in relative stimulus duration results in the longer-duration stimulus extending even farther in time beyond the shorter duration stimulus. In this case, any influence on onset-order discrimination would likely occur via a backward masking mechanism. In the SOFA experiment, as the difference between the two stimulus durations increases, the longer duration stimulus tends to extend earlier in time relative to the shorter duration stimulus, in which case, the effect of forward masking on offset-order discrimination might be expected to increase. Thus, the selective effect of stimulus duration discrepancy on offset-order sensitivity suggests that increased stimulus duration may elicit forward masking more effectively than backward masking in the Pacinian system. This explanation is consistent with the notion of a temporal asymmetry of vibrotactile forward and backward masking, which we have proposed to underlie the tendency for offset-order discrimination thresholds to exceed those for onset-order discrimination.

Just as we have observed, Yuan et al. (2005a) found that onset-order discrimination was substantially affected by the relative levels of paired stimuli, whereas the effect of roving stimulus durations was nominal. As in the current study, Yuan and colleagues adjusted the amplitudes of the two shortest-duration stimuli (see Methods), and they postulated that these amplitude adjustments may have contributed to the negligible effects of roving stimulus duration on discrimination thresholds. This seems like a reasonable explanation, in light of the amplitude-dependence of thresholds in the onset-order discrimination experiment. The upper left-hand plot in Figure 3-6 indicates that, particularly for the lowest $|SOA|$ value examined, performance is better when the amplitude of the earlier-onset stimulus is larger relative to the amplitude of the later-onset stimulus (towards the right on the abscissa). Now consider the likely effects on the upper right-hand plot of Figure 3-6 when we raise the amplitudes of all

50-ms and 100-ms stimuli (by 6 dB and 3 dB, respectively). Specifically, in duration category 1, the brief stimuli with increased amplitudes will always have the earlier onset, and in duration category 7, they will always have the later onset. Thus, the amplitude adjustment would tend to improve performance in category 1 trials, in which the earlier onset stimulus would otherwise be less salient. In category 7 trials, the adjustment would tend to undermine performance, in which the longer, earlier-onset stimuli might otherwise mask the shorter stimulus more effectively.

The same amplitude adjustments are applied to the two shortest-duration stimuli in the offset-order discrimination experiment, in which the effects of roving stimulus duration on thresholds are quite substantial. In this case, it is likely that the adjustments made to 50-ms and 100-ms stimulus amplitudes have favored the apparent dependence of performance on relative stimulus durations. In the bottom left-hand plot in Figure 3-6, we have noted that subjects perform increasingly well as the amplitude of the later-offset stimulus increases relative to the earlier-offset stimulus (towards the right on the abscissa). In the bottom right-hand plot, the brief, amplitude-adjusted stimuli will always have the later offset in duration category 1 trials and the earlier offset in duration category 7 trials. As a result, subjects most likely perform better in category 1 trials and worse in category 7 trials than they would without the amplitude adjustment.

3.4.4 Implications for tactile displays of speech

A significant motivation for the present study was to guide the further development of tactual speech aids for the deaf. In particular, discriminating between voiced and unvoiced consonants (e.g., /b/ and /p/) poses a major challenge to deaf individuals when communicating via lipreading. Yuan et al. (2004a) developed an acoustic cue comparable to voice-onset timing (VOT), which they have called "Envelope-onset asynchrony" (EOA). The EOA cue is derived from the difference in onset-timing between the envelopes of a 3000 Hz high-pass filtered band and a 350 Hz low-pass filtered band (Yuan et al., 2004a). Analyzing isolated, spoken pairs of CVCs with initial consonants that differ only in the feature voicing, they showed that the amplitude envelopes of the two filtered bands for a voiced initial consonant tend to have similar onsets, separated by several tens of milliseconds at most, whereas for unvoiced initial consonants, the onset of the high frequency envelope generally precedes that of the low frequency envelope by considerably longer durations. EOAs are consistently on the order of 50 to 200 ms larger for an unvoiced initial consonant than for its voiced counterpart. A comparable relationship between the voicing of final consonants and the envelope-offset asynchrony (EOFA) has also been observed (Yuan et al., 2004b; Yuan

and Reed, 2005). EOFA values for paired voiced and unvoiced final consonants typically differ by 200 to 300 ms.

Yuan and colleagues further established that normal-hearing individuals could exploit vibrotactile presentation of amplitude envelope information from the two speech bands to discriminate EOA and EOFA cues, and thereby distinguish between voiced-unvoiced consonant pairs in both initial (Yuan et al., 2005b) and final (Yuan et al., 2004b; Yuan and Reed, 2005) positions.

The present study sought to evaluate possible limits of cross-channel tactual temporal processing among pre-lingually deaf individuals, in order to determine the need for corresponding constraints on the tactual encoding of acoustic signals by a tactile speech aid. The results suggest that most congenitally deaf participants in this study should have sufficient temporal resolution to take advantage of tactually-presented EOA and EOFA cues to supplement lipreading. Moreover, the powerful influence of relative stimulus levels on temporal order perception across skin loci in both the SOA and SOFA experiments has underscored the importance of amplitude range compression in the vibrotactile transduction of an acoustic speech signal. More generally, the present study contributes fundamentally to our understanding of sensory processing in persons with early-onset deafness.

The study described in the next two chapters examines the abilities of deaf individuals to discriminate the voicing distinction in speech sounds using a modified version of the tactual EOA/EOFA scheme, which incorporates additional articulatory cues with the objective of establishing a more comprehensive tactual speech representation.

3.5 Summary

- Tactual detection thresholds are comparable among congenitally deaf and normal-hearing adults for sinusoidal stimuli ranging in frequency between 2-300 Hz.
- Tactual onset-order discrimination thresholds for congenitally deaf and normal-hearing subjects averaged 83 ms and 58 ms, respectively; the difference is not statistically significant.
- Tactual offset-order discrimination thresholds for congenitally deaf and normal-hearing subjects averaged 139 ms and 121 ms, respectively; the difference is not statistically significant.
- Tactual offset-order discrimination thresholds were, on average, roughly 1.8 times larger than tactual onset-order discrimination threshold, unlike in audition, where offset-order thresholds have been found to be comparable to, if not lower than, onset-order thresholds.
- Level differences between paired stimuli play a consistent role in the tactual discrimination of both onset-order and offset-order, for subjects in both groups.
- Duration differences play a consistent role in the tactual discrimination of offset-order, but not onset-order, for subjects in both groups.
- Discrimination thresholds for tactual offset-order are generally larger than those for tactual onset-order, and individual subjects' performance on the two tasks is correlated ($r = 0.852$).
- Tactual temporal resolution among congenitally deaf subjects should be sufficient for utilizing a tactually-encoded temporal speech cue, for displaying crucial voicing information that is not available through lipreading.

Acknowledgments

The work described in this section has been published as:

Moallem TM, Reed CM, Braida LD. (2010). "Measures of tactual detection and temporal order resolution in congenitally deaf and normal-hearing adults," *J. Acoust. Soc. Am.* 127(6):3696-709.

This research was supported by grants from the National Institutes of Health (Grant Nos. 5T32 DC000038, R01-DC000117, and R01-DC00126). The authors would like to thank Ruth Litovsky and two anonymous reviewers for their helpful suggestions in revising an earlier version of the manuscript.

Chapter 4.

Development and Implementation of a Real-time Tactual Speech Display

4.1 A Tactual Cue for Voicing

Yuan et al. (2004a) described an acoustic cue comparable to VOT, which they have called "Envelope-onset asynchrony" (EOA). The EOA cue is derived from the difference in onset-timing between the envelopes of a 3000 Hz high-pass filtered band and a 350 Hz low-pass filtered band (Yuan et al., 2004a). Analyzing isolated, spoken pairs of CVCs with initial consonants that differ only in the [\pm voiced] feature, they showed that the amplitude envelopes of the two filtered bands for a voiced initial consonant tend to have similar onset timing, separated by several tens of milliseconds at most, whereas for unvoiced initial consonants, the onset of the high frequency envelope generally precedes that of the low frequency envelope by considerably longer. EOAs are consistently on the order of 50 to 200 ms larger for an unvoiced initial consonant than for its voiced counterpart. Such is the case for all eight pairs of English consonants that have the same place and manner of articulation, but differ in voicing. A comparable relationship between the voicing of final consonants and the envelope-offset asynchrony (EOFA) has also been observed (Yuan et al., 2004b; Yuan and Reed, 2005). EOFA values for paired voiced and unvoiced final consonants typically differ by 200 to 300 ms.

Yuan and colleagues next sought to determine whether or not normal hearing individuals could exploit vibrotactile presentation of amplitude envelope information from the two bands to discriminate EOA and EOFA cues, and thereby distinguish between voiced-unvoiced consonant pairs. Starting with video recordings of two female speakers producing isolated CVCs, Yuan et al. (2005b) separated the audio into two channels. One channel was high-pass filtered with a cutoff frequency of 3000 Hz, and the other channel was low-pass filtered with a cutoff frequency of 350 Hz. The envelope of each of these signals was extracted by rectification and low-pass filtering at 25 Hz. Amplitude envelope values below a predetermined threshold level were set to zero, in order to eliminate random noise fluctuations. The envelopes of the two filtered signals were then used to modulate sinusoidal carrier signals driving two

tactual stimulators. The stimulation frequencies were 50 Hz at the left thumb and 250 Hz at the left index finger. Supra-threshold values of the envelope produced proportional increases in carrier amplitudes, which were otherwise set to drive the stimulators at the average detection threshold level of the carrier frequency.

Results indicated that providing subjects with this tactual information during lipreading of isolated CVCs facilitated identification of voiced and unvoiced consonants in both initial (Yuan et al., 2005b) and final (Yuan et al., 2004b; Yuan and Reed, 2005) positions. For example, using a two-interval, two-alternative forced choice (2I-2AFC) procedure employing CVC nonsense syllable stimuli, Yuan and colleagues (2005b) examined the ability of four normal-hearing subjects to discriminate initial consonant voicing for each of the eight pairs of English consonants that differ only in the voicing feature. In each trial, the initial consonants of the two CVCs presented differed only in the voicing feature, and subjects were asked to indicate the order in which the CVCs with voiced and voiceless initial consonants were presented. Figure 4-1 (reproduced from that study) shows average values of d' across subjects for each consonant pair under each of three testing conditions: lipreading-alone (L), tactual-alone (T), lipreading and tactual together (L+T). As one might expect due to the lack of visible cues for voicing on the face of a speaker, performance in the lipreading-alone condition (L) falls close to chance levels (i.e., $d'=0$) across all consonant voicing pairs. By contrast, subjects' ability to make voicing distinctions improves substantially across all consonant pairs under the two conditions that incorporate the tactual EOA cue (T and L+T).

The capacity of profoundly deaf individuals with various etiologies to discriminate the tactual cues of the EOA/EOFA scheme remains unclear. Thus, an important motivation for the current research is to examine the usefulness of this type of acoustic-tactile transduction approach for a profoundly deaf population. Moreover, we seek to extend the articulatory information content of the EOA/EOFA encoding scheme, to convey more detailed correlates of articulatory manner, place, and vocal periodicity, which are ambiguous during visual speechreading.

4.2 Practical Objectives and Constraints

In developing an effective sensory prosthesis, we wish to incorporate certain assistive functionalities, while orienting and constraining our efforts in accordance with practical issues surrounding real-world implementation. These include feasibility of low-cost production, widespread distribution, usability, wearability, device maintenance, and training requirements. Of course, these issues need not all be

addressed in at this stage of implementation. However, by designing and implementing the technology with these constraints in mind, we hope to maximize the potential for impact on a global scale.

4.2.1 Fine Temporal and Gross Spectral Enhancement

To the highly effective consonant voicing discriminability demonstrated by Yuan et al. (2004b, 2005b), we wish to add three primary functional enhancements. First, instead of fixed frequency vibrotactile carrier modulation, the signal on each stimulator channel should directly reflect the periodicity and/or aperiodicity of its corresponding band-filtered acoustic signal. Second, the encoding algorithm for each channel should retain and convey fluctuations in the band-filtered acoustic signal amplitude that fall

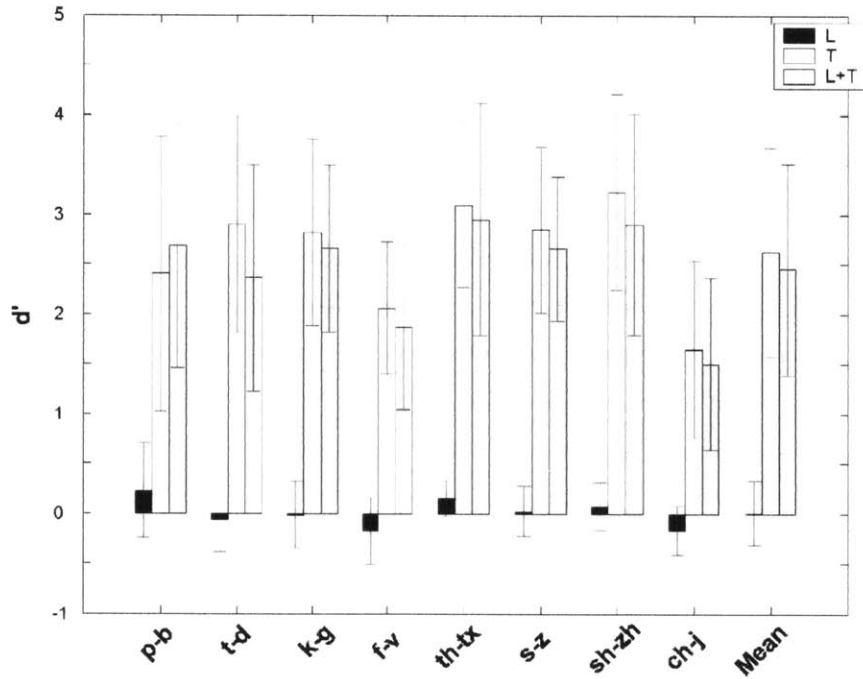


Figure 4-1. Results of 2I-2AFC initial consonant voicing discrimination testing, performed by four normal-hearing subjects (Yuan et al. 2005b). In each trial, initial consonants of two CVC stimuli differed in voicing feature only, and subjects indicated the order in which the CVCs with voiced and voiceless initial consonants were presented. Values of d' are averaged across subjects for each of eight English consonant pairs, under three testing conditions: lipreading-alone (L), tactual-alone (T), lipreading and tactual together (L+T). [Figure reproduced from Yuan et al., 2005b]

within the frequency range of tactual sensitivity (much of which Yuan et al. eliminated by applying "smoothing" filters). Finally, in addition to the high and low-frequency channels implemented by Yuan et al. (2005b), we will add a mid-frequency channel, contacting a third skin locus, in the interest of affording the user a more complete (albeit crude) representation of formant progressions and articulatory place.

4.2.2 Suitability for Low-Cost Mobile Implementation

Experiments carried out in support of this thesis have been restricted to a laboratory setting, using an assortment of non-mobile computing, signal processing, and tactual display equipment. Although laboratory implementation places relatively few inherent limitations on processing speeds, mechanical complexity, energy expenditure, or equipment costs, we do best to adopt constraints that are consistent with the dimensions and ultimate purchase/maintenance cost of the desired product.

It is crucial then to recognize that the vast majority of severely and profoundly deaf individuals, the potential beneficiaries of this tactual speech technology, reside in low- and middle-income countries (World Health Organization, 2006). As such, we strive to develop a simple and versatile tactual speech strategy, amenable to implementation on a variety of common, low-cost mobile platforms, utilizing minimal computing resources. Unfortunately, a suitable low-cost tactile (actuator) interface is not commercially available at present. However, a wide variety of vibrotactile technologies can be implemented simply and cheaply, and it seems reasonable to expect that the challenge of either identifying or developing a suitably low-cost, mobile tactual interface is surmountable. The present author is committed to addressing the hardware issue as the next phase of this project. The remaining portion of the present dissertation focuses on the development and evaluation of tactual speech encoding and training strategies amenable to implementation on a low-cost, microprocessor-based mobile platform.

4.3 Tactual Apparatus

Tactual stimulation at the finger pads is achieved using a Tactuator multi-finger display, the same apparatus used in Chapter 3 of this document, which is described in detail in Tan and Rabinowitz (1996) and Yuan (2003). The Tactuator user interface consists of three rods that lightly contact the thumb, index finger, and middle finger of the left hand, in an arrangement that allows for a natural hand configuration

(illustrated in Figure 4-2). Each rod is moved independently by a head-positioning motor, under the control of an external voltage. This driving voltage is regulated by a feedback control mechanism, which compares a control signal with the output of an angular position sensor that is coupled to the rotation axis. The Tactuator is designed to deliver stimuli with frequencies from dc to around 400 Hz, spanning most of the cumulative receptive range of the four major types of somatosensory afferent receptors, which together mediate kinesthetic, flutter, vibratory, and cutaneous tactual percepts (Bolanowski et al., 1988). Each rod has roughly a 26 mm range of motion, meaning that it can reliably deliver stimuli at levels between detection threshold and about 50 dB above threshold, throughout its useful frequency range. Within the scope of these parameters, the displacement of each stimulator is proportional to the time-varying amplitude of its control signal waveform. The effect of loading due to light finger contact is minimal (Tan and Rabinowitz, 1996).

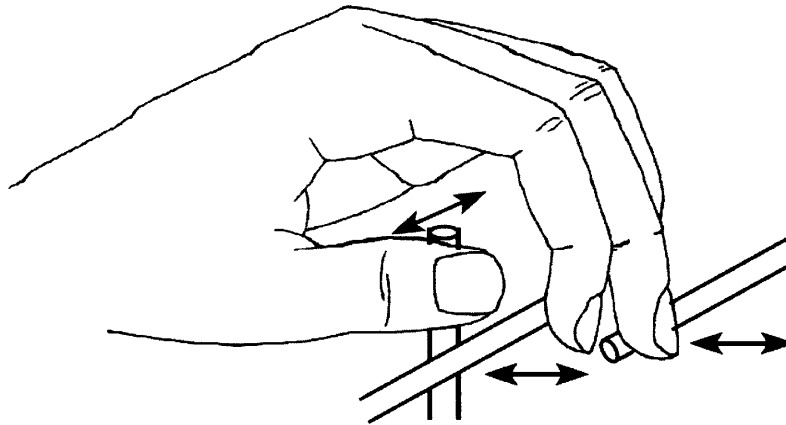


Figure 4-2. Schematic illustration of the Tactuator three-finger interface configuration.
[Figure reproduced from Tan and Rabinowitz, 1996]

4.4 Tactual Speech Transduction Scheme

The tactual transduction scheme developed for this thesis is illustrated in Figure 4-3. Following analog-to-digital conversion, the digitized audio signal is passed through a first-order Butterworth high-pass filter with a 2 Hz cutoff frequency in order to eliminate DC content. The signal is then split into three separate streams, each of which is filtered to attenuate content outside of one of the following frequency bands:

50-400 Hz, 800-2200 Hz, and 3000-8000 Hz. In each case, this is achieved using sequential fourth-order low-pass and high-pass Butterworth filters.

The signal corresponding to the lowest frequency band (50-400 Hz) is half-wave rectified and then passed through an additional fourth-order Butterworth 400 Hz low-pass filter, after which sample amplitudes are adjusted according to the μ -law logarithmic compression algorithm (Stremmer, 1990), as a means of compensating for the substantially reduced dynamic range of vibrotactile sensitivity relative to audition. Subsequent tactual equalization entails application of a first-order low-pass filter with cutoff frequency of 20 Hz, which effectively adjusts the spectral weighting of the signal to compensate for the decline in tactual sensitivity with decreasing frequency below about 200 Hz. The resulting signal is converted to an analog voltage and fed to the PID feedback controller input corresponding to the middle finger stimulator channel of the Tactuator.

Unlike the lowest frequency band signal, the two higher frequency band-pass filtered signals have spectral content that lies primarily outside of the frequency range of tactual sensitivity. Rectification of these two signals has the effect of introducing low-frequency (tactually-discernable) content, corresponding to the amplitude modulation contours of their high frequency content. For reasons discussed below, full-wave (rather than half-wave) rectification is applied to these signals. Following rectification, both signals are passed through a fourth-order Butterworth 400 Hz low-pass filter. For the rectified mid- and high-frequency signals, which carry substantial high-frequency content, application of the 400 Hz low-pass filter is particularly essential to limit the signals primarily to the frequency range of tactual sensitivity, while retaining quasi-periodic and aperiodic content spanning the typical range of vocal fundamental frequencies. After low-pass filtering, the signals are sequentially subjected to amplitude (μ -law) compression, tactual equalization, and conversion to analog voltage signals. The mid-frequency and high-frequency signals are then fed to the PID controllers corresponding to the Tactuator's thumb and index finger channels, respectively.

The rectification step in this transduction scheme was initially implemented as a half-wave rectification, in accordance with a simple model of cochlear transduction (Lazzaro and Mead, 1989). For the highest frequency band, the choice of half-wave vs. full-wave rectification has only minimal impact on the output waveform. For the mid-frequency band, the rectification type can make a substantial difference in the spectral composition of the output signal, most notably during periods of sustained vocalization. A voiced speech waveform reflects the polarity of the underlying glottal source, and the amplitudes of the largest positive and negative peaks in each fundamental period can differ substantially. As a result, the

polarity of the digitized speech signal at the time of half-wave rectification (or equivalently, the decision to zero either positive or negative values of the signal) can substantially influence the relative contributions of the fundamental frequency and its lowest harmonics to the output signal. Given that the microphone configuration of a tactile aid remains unchanged, the polarity of a speech signal detected directly (in the presence of a live speaker) will also be consistent. However, because human audition is generally insensitive to speech signal inversions, the polarity of speech conveyed via electronic media varies on a case-by-case basis. Moreover, speech signal polarity is often ambiguous (as in cases of noise contamination, variable source position/orientation, and reconstitution following encoding for transmission or storage). Synthesized speech, which is increasingly prevalent in a variety of social, commercial, and educational contexts, need not have any conventional polarity. The practical perceptual significance of differences in signal polarity when using half-wave rectification in the present speech transduction scheme remains to be determined.

The use of half-wave rectification would allow application of the same processing scheme across all channels, varying only the band-pass filter parameters. However, for purposes of the present study, the use of full-wave rectification in the mid- and high-frequency channels only (as depicted in Figure 4-3) provides for a straightforward and effective transduction strategy that is insensitive to speech signal polarity. Half-wave rectification is used for the low-frequency channel only, as full-wave rectification could potentially obscure the fundamental frequency of voicing.

The speech encoding strategy outlined above incorporates key elements of the aforementioned spatio-temporal display of consonant voicing cues (Yuan et al., 2005b) and provides further temporal and spectral information, intended to support the discrimination of articulatory manner and gross formant structure.

Yuan and colleagues used a two-channel system, in which the amplitude envelopes derived from a high-pass filtered channel (3000 Hz cutoff) and a low-pass filtered channel (350 Hz cutoff) were used to modulate sinusoidal vibrations at the left index finger and thumb, respectively. In a series of studies, they demonstrated that unvoiced consonants in both syllable-initial and final positions are consistently associated with substantially larger inter-channel asynchronies, which are effectively discriminated by trained, normal-hearing subjects (Yuan et al., 2004a, 2004b, 2005b; Yuan and Reed, 2005). However, their approach of

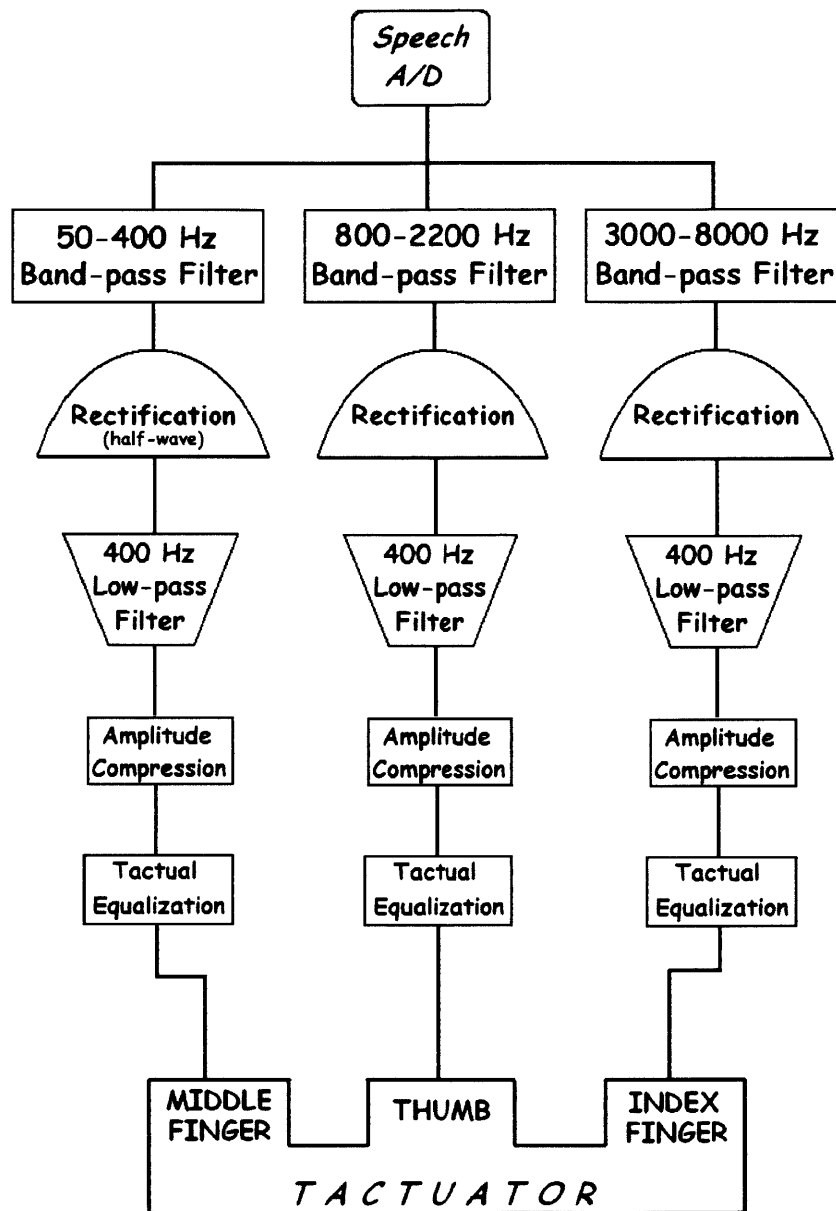


Figure 4-3. Flowchart illustration of vibrotactile speech transduction scheme. A speech signal, from an amplified microphone or alternative audio source, enters soundcard via a low-latency (ASIO) input for analog-to-digital conversion. DC content is eliminated from the digitized signal using a first-order 2 Hz high-pass filter. The signal is then split into three streams, each of which is band-pass filtered to attenuate content outside of the specified frequency bands and further processed as shown (see text for complete description). The three voltage signals derived from the 50-400 Hz, 800-2200 Hz, and 3000-8000 Hz bands serve as control inputs to the middle finger, index finger, and thumb channels of the Tactuator apparatus, respectively.

amplitude modulating fixed-frequency sinusoidal carriers has the drawback of discarding low frequency modulations inherent to speech, including those reflecting the fundamental frequency (F0) of voicing. Although the output of the 350 Hz low-pass channel correlates strongly with voicing amplitude, it does not reflect the frequency range or contour of F0, nor does it allow for discrimination between voicing and aperiodic noise.

In auditory speech perception, F0 variations contribute substantially to prosodic structure, phonetic discrimination, and stream segregation, all of which take on increasing significance as the acoustic signal-to-noise ratio decreases. In particular, voicing F0 facilitates speech discrimination in the presence of multiple talkers or speech-like noise (Brokx and Nootboom, 1982; Bird and Darwin, 1998; Assmann, 1999). Prosodic cues, consisting largely of variations in voicing frequency and amplitude, are instrumental to speech comprehension even at a high signal-to-noise ratio (SNR), and a speaker will often exaggerate such cues further as background noise levels increase. In fact, much information can be communicated through stresses and intonation patterns alone, or in combination with highly degraded or modified phonetic cues (Blessner, 1969).

During voicing, the fundamental frequency is evident across the frequency spectrum — not only is it reflected in the harmonic content of the acoustic signal, but the amplitude of each harmonic component is modulated at the fundamental frequency. As illustrated in Figure 4-4, these cross-spectrum modulations are strongly represented in the activity patterns of auditory nerve (AN) fibers over a vast range of characteristic frequencies. Thus, even when confronted with low frequency noise or the attenuation of low-frequency content (as when communicating from a distance), the AN population code may still clearly reflect the voicing F0 and variation thereof. The spectrum-wide contribution of voicing to an acoustic signal can be identified on the basis of amplitude modulation that shares the period of the fundamental. Moreover, the frequencies over which these modulations vary fall (conveniently) within the range of tactual sensitivity.

With this in mind, we sought to implement a vibrotactile transduction scheme that conveys the time-varying amplitude of each frequency-filtered speech band, retaining the voicing F0 and other low-frequency modulations while preserving the cross-channel timing information required for consonant voicing discrimination. The frequency content of the 50-400 Hz band, including the vocal F0, falls naturally within the range of vibrotactile sensitivity. Subsequent to signal rectification, F0 amplitude modulations of higher vocal harmonics predominant in the mid- and high-frequency bands are reflected spectrally, within the frequency range of tactual sensitivity.

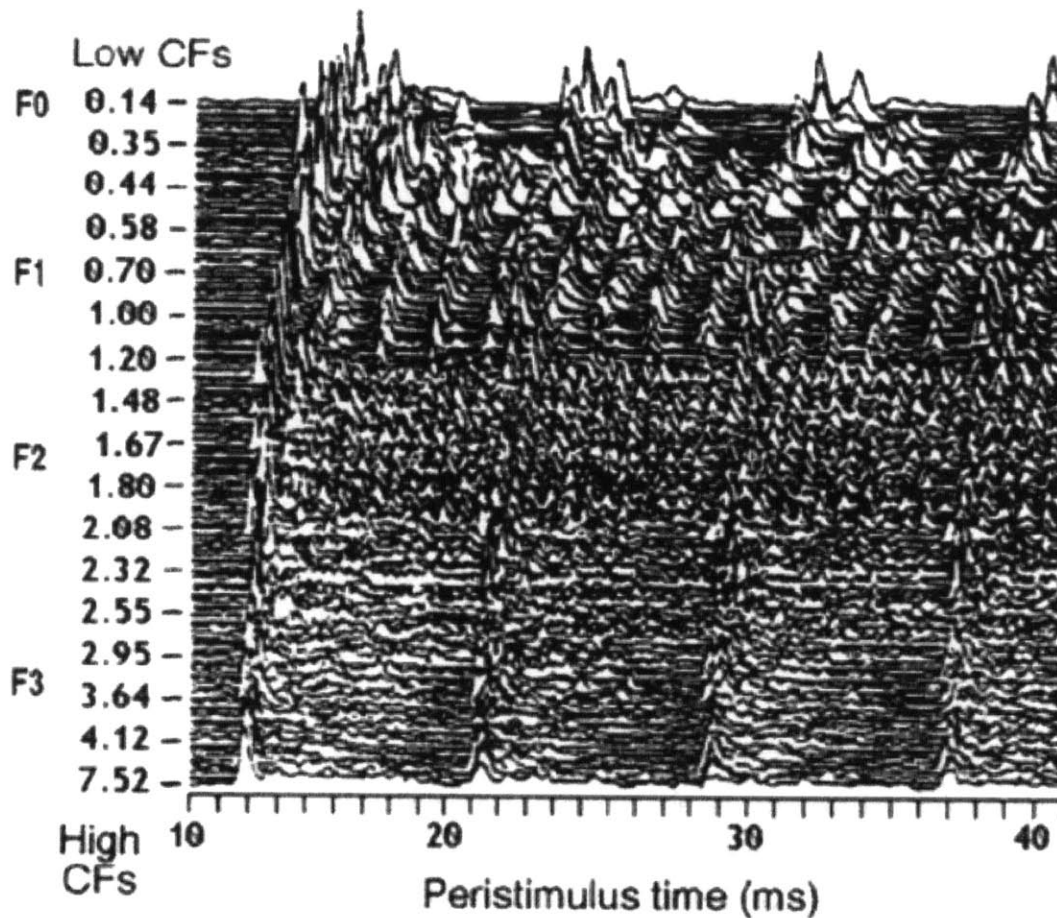


Figure 4-4. Estimated firing rates of 178 AN units, averaged over multiple presentations of the syllable /da/. The period of the fundamental frequency of voicing (approximately 8 ms) is clearly represented in the firing rates of AN fibers with CFs spanning the depicted range of approximately 140 to 7500 Hz. The temporal structure of each unit response tends toward one of several distinct activity patterns, reflecting entrainment to the harmonic components that dominate their respective formant regions (F1, F2, F3). Those fibers with the lowest CFs appear to entrain to the fundamental itself (F0). [Reproduced from Secker-Walker and Searle (1990), who analyzed data from Young and Sachs (1979)]

Interestingly, if the stimulator corresponding to the lowest frequency speech band is shifted to a more proximal position on the hand (just below the base of the finger), at a similar distance from each of the other finger contacts, the spatial distribution of tactual stimulators delineates a more contiguous, two-dimensional contour along the volar surface, which one might interpret perceptually as the surface of an object, gripped lightly in the hand. Tactual presentation of speech stimuli can further enhance this

percept, providing that one permits the illusory object surface to take on a dynamic quality, shifting its position one way and then the other, bubbling outward or bursting locally and then expanding more diffusely. A similar illusory surface percept, albeit less pronounced, can also be achieved with the current tactual stimulator configuration. Certain speech sounds, the alveolar consonants and the vowel /u/ in particular, elicit consistent, well-localized sensations that are relatively independent of phonetic context. However, speech sounds with less distinctive spectral characteristics, which are not consistently associated with a localized percept or spatiotemporal pattern, may still prove discernible on the basis of spectrally diffuse temporal structure. For example, when uttered in syllable-initial position, the unvoiced plosive /p/ often elicits a faint pulse-like sensation. Unless the plosive burst is particularly well enunciated, the percept is typically not well localized, but might rather be characterized perceptually as an impulse transmitted to the contact surface as a whole, as though resulting from a mild impact to the underside of the illusory, lightly-gripped object. This pulse-like percept is invariably absent when the voiced plosive /b/ is uttered in a comparable phonetic context.

The next section provides a detailed account of the correspondence between key articulatory features, the encoded three-channel vibrotactile signal, and the resulting tactual percepts, which in turn provide a robust basis for reception of phonetic information, particularly when combined with visual speechreading. The following chapter then describes an experimental study carried out to assess the practicability of this vibrotactile articulatory encoding scheme.

4.5 Vibrotactile Encoding of Articulatory Features

4.5.1 Vowels

Figure 4-5 presents the three-channel vibrotactile (displacement) signals corresponding to segments of the English vowels /i/, /u/, and /a/. The fundamental frequency of voicing (F_0) is clearly reflected in the time waveforms of all channels. The bottom waveform (50-400 Hz) corresponds to the signal presented to the middle finger, the middle waveform (800-2200 Hz) corresponds to the signal presented to the thumb, and the top waveform corresponds to the signal presented to the index finger. Upon close inspection, it may be observed that the peaks in corresponding F_0 periods reveal a relative phase shift of several milliseconds among the three waveforms, reflecting the distinct phase characteristics of the high and low frequency filtering processes. Cochlear filtering produces similar phase disparities across frequency channels, which is evident in the response latencies of AN units with differing characteristic frequencies.

This phenomenon is well illustrated in Figure 4-4, where AN fiber response latencies are observed to decrease from 14 - 15 ms at the lowest CFs shown to approximately 12 ms at the highest CFs.

4.5.2 Syllable-Initial Consonants

Figure 4-6 depicts the three-channel waveforms for vibrotactile signals corresponding to 12 recorded CVC utterances. Each CVC includes a particular consonant in both initial and final position, separated by the vowel /i/. All 12 consonants used in this study are represented, and consonants differing in voicing only ("voicing pairs") are positioned adjacently. Due largely to tactual equalization (implemented as a first-order low-pass filter as described above), the low-frequency content (on the order of 1-20 Hz) of all three vibrotactile signals exhibits much larger amplitudes than higher frequency content, in the range of voicing F0. (In Figure 4-6, the large, rolling peaks reflect low-frequency content, whereas the dark regions atop these peaks reflect the smaller amplitude high frequency content, much of which is

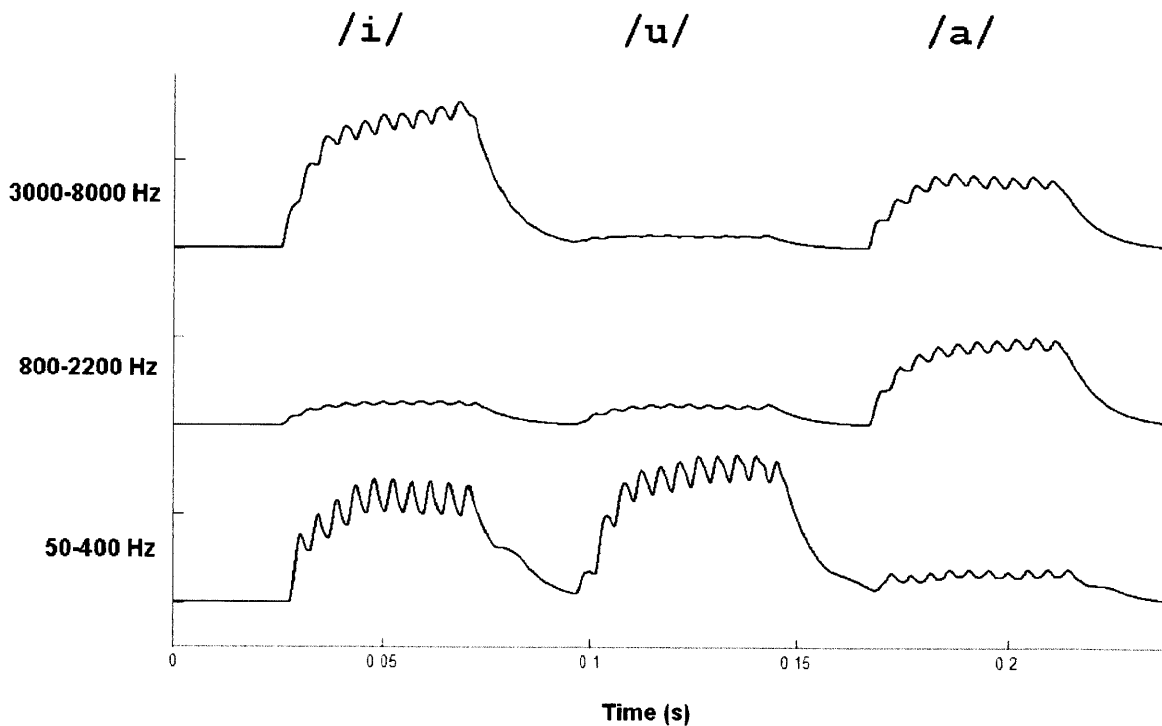


Figure 4-5. Three-channel vibrotactile signals corresponding to brief segments of the vowels /i/ in "beed", /u/ in "zood", and /a/ in "gahd" produced by an adult female (bordered on each side by 25 ms silence). Labels on the left indicate the frequency band from which each speech signal derives.

contributed by voicing.) The disproportionately large amplitudes of low frequency signal components compensate for the disproportionately greater sensitivity of the human tactual system to higher frequency signal components. Thus, on the 50-400 Hz channel (which contacts the middle finger), despite the fact that the low-frequency component amplitudes appear quite large in Figure 4-5, the high frequency vibratory component is typically more prominent in terms of sensation level.

To avoid confusion between references to low/high frequency content of the vibrotactile signals and low/middle/high (acoustic) frequency channels, the latter are delineated as the *M*, *T*, and *I* channels, as indicated to the right of each waveform in Figure 4-6. These labels reference the middle finger, thumb, and index finger, respectively, to which the signals were delivered in this study.

Two primary factors account for the relative absence of activity on the T-channel in Figure 4-6. First, the vowel /i/, present in all 12 CVC utterances, is predominantly characterized by a low first formant and high second formant, as is clearly reflected in the leftmost traces of Figure 4-5. The T-channel serves to distinguish open (or low) vowels such as /a/ and /æ/ from adjacent consonants and close (or high) vowels. Although open and close vowels appear visually distinct on the face of a speaker, activity on the T-channel may help to delineate consonant-vowel transitions and provide a tactile context for recognition of consonant cues that vary as a function of coarticulatory context. Second, the utterances represented in Figure 4-6 were all produced by a single adult male speaker. By contrast, adult female speakers tend to exhibit more T-channel (and less M-channel) activity during the glottal burst consonants /g/ and /k/, as illustrated by the vibrotactile signal for the utterance /k u dʒ/ in Figure 4-7 (left). Thus, for some speakers, the T-channel provides an effective cue for the articulatory place of glottal consonants. (The T-channel's role in facilitating discrimination of approximants (e.g., /l/, /ɹ/, /j/) and diphthong vowels has yet to be examined.)

The most consistent and robust cues for consonant voicing are the relative onset and offset timings of the signals on the *I*-channel and *M*-channel. In the case of unvoiced initial consonants, activity on the *I*-channel typically precedes that on the *M*-channel by anywhere from about 50 ms to more than 200 ms. At syllable onset, the signal on the *I*-channel usually consists of a low-frequency, high-amplitude component and an aperiodic, high-frequency, low-amplitude component. The low-frequency component manifests as a variably-sized displacement at the index finger, the percept of which tends to dominate over the concurrent high-frequency vibration. For most unvoiced initial consonants, the *M*-channel is

quiet until the onset of the vowel, which produces a smooth vibratory percept reflecting the vocal F0. The vibrotactile waveforms corresponding to the unvoiced consonant /s/ in initial position (e.g., Figure 4-6, top left) effectively illustrate this cross-channel activity pattern.

The vibrotactile patterns observed for the initial unvoiced consonants /s/, /t/, /f/, /tʃ/, /p/ and /k/ are all similar in that the I-channel dominates the onset portion of each signal. However, the frequency content and duration of the I-channel signal can differ substantially, and in some cases, the aperiodic onset burst is also reflected on either or both of the other channels. These signature differences in the vibrotactile onset pattern should provide a strong perceptual basis for distinguishing articulatory manner, and in some cases, articulatory place.

Via lipreading alone, /s/ and /t/ are virtually indistinguishable from one another, in both syllable-initial and final positions. The difference in the I-channel onset patterns of /s/ and /t/ (Figure 4-6, top left and middle left, respectively) thus could offer a particularly powerful vibrotactile cue for discriminating articulatory manner. The relative amplitudes of the two I-channel onset peaks are not invariant features, and do not provide a consistent basis for discrimination. However, close inspection of the two I-channel signals reveals that the slope of the /t/ onset is substantially greater than that of the /s/ onset. Also, regardless of relative amplitude, the /t/ signal is generally transient in nature, while the /s/ peak exhibits more of a plateau. In both initial and final position, these differences in onset slope and duration offer a very reliable basis for distinguishing the plosive /t/ from the fricative /s/. The percept associated with the sharp sloping /t/ signal is abrupt and concussive, whereas an /s/ produces a gentler, more sustained fluttering sensation.

For the labial consonant /p/ in initial position, the onset burst on the I-channel is typically accompanied by a brief impulse on the M-channel, which produces a punctate deflection at the middle finger. The combined I-channel and M-channel activity during an initial /p/ result in an often non-localized tap sensation leading into the vowel. This quality of a somewhat "rough" onset effectively distinguishes the initial /p/ percept from that of a voiced initial /b/.

For the unvoiced glottal consonant /k/ in initial position, the I-channel onset is generally accompanied by a small burst of activity on the M-channel, the T-channel, or both, depending on the speaker and phonetic context. The resulting deflection at the middle finger and/or thumb, in conjunction with the large deflection at the index finger, provides a cue for articulatory place, distinguishing glottal /k/ from alveolar /t/, which are often confused in exclusively visual speechreading.

Voiced initial consonants are characterized by I-channel and M-channel onsets that are more nearly concurrent, usually separated by less than 50 ms. For example, the waveforms for the voiced consonant /z/ in initial position (Figure 4-6, top, second from left) demonstrate approximately simultaneous I-channel and M-channel onsets. The I-channel exhibits a low-frequency, high-amplitude onset peak, similar to that of the initial /s/, which in both cases corresponds to frication noise. However, unlike the /s/ vibrotactile signal (which consists of frication only), the /z/ signal includes the periodic low-amplitude vibration characteristic of the voicing F0, on both the I-channel and M-channel, beginning at the initial release of the consonant.

The vibrotactile patterns observed for the initial voiced consonants /z/, /d/, /v/, /dʒ/, /b/ and /g/ are all similar in that the onsets of I-channel and M-channel activity occur close in time, usually separated by less than 50 ms. In cases of consonant pre-voicing, vibration on the M-channel has an earlier onset, as is observed for the strongly pre-voiced initial /b/ shown on the right of Figure 4-7. For the initial consonant /v/ (Figure 4-6, middle right), I-channel and M-channel signal onsets are nearly simultaneous, yet due to the difference in onset amplitude between the two channels, the F0 vibration percept at the middle finger distinctly precedes the index finger percept, thus clearly distinguishing the voiced /v/ from its unvoiced counterpart /f/, which shares the same viseme. By contrast, the I-channel onset of an initial /dʒ/ generally precedes the M-channel onset by at least 40-50 ms, and the sharp transient character of the I-channel peak ensures that the onset asynchrony is clearly perceived. Thus, the tactual pattern of an initial /dʒ/ is perhaps most likely among the voiced English consonants to be confused for an unvoiced consonant. However, when articulatory place and manner are recognized, an initial /dʒ/ is readily distinguished from its unvoiced counterpart, /tʃ/, for which the I-channel onset commonly precedes the M-channel onset by 200-300 ms.

As was observed for the corresponding unvoiced consonants, the frequency content and duration of the I-channel signal differs substantially among the voiced initial consonants represented in Figure 4-6. These differences, in conjunction with the early onset of voicing reflected on the M-channel, provide the requisite perceptual cues for distinguishing consonant voicing and articulatory manner.

While the vibrotactile signals do not identify articulatory place unambiguously, they can provide useful constraints, ruling out certain place confusions, and thus increasing the efficacy of information reception

overall, or the likelihood of successful communication in any given instance. For example, during lipreading, a slight lapse in visual attention can easily result in confusion between /b/ and /d/ in syllable-initial position. Looking at the waveforms for /b i b/ and /d i d/ in Figure 4-6, one readily observes the sharp onset peak on the I-channel for the consonant /d/, the tactual percept of which serves to disambiguate the phoneme, ruling out a confusion with /b/. Similarly, with the benefit of tactual cuing, confusions between /v/ and /z/ are highly unlikely, even when the visual percept is disrupted.

For the voiced initial glottal consonant /g/, the aperiodic onset burst is often reflected strongly enough on either the T-channel, the M-channel, or both, such that it provides a valuable cue for articulatory place, tactually distinguishing the initial /g/ from the voiced alveolar consonant /d/. Such is the case for the initial consonant in the CVC /g i g/, depicted in Figure 4-6. Of course, this cue is in many instances so subtle as to escape detection under the current transduction scheme. Although it may facilitate place identification in some instances, it is not an invariant place cue.

4.5.3 Syllable-Final Consonants

The vibrotactile representations of the 12 consonants in syllable-final position differ from their syllable-initial counterparts in two key respects. First of all, the onset asynchrony observed between the I-channel and M-channel for unvoiced initial consonants is temporally reversed for the corresponding final consonants. In other words, the offset of activity on the M-channel precedes the offset of I-channel activity. Moreover, the temporal asynchrony is generally substantially larger for a given unvoiced consonant at the end of a syllable than at the beginning. Activity on the I-channel typically persists anywhere from about 100 ms to more than 300 ms beyond the offset of M-channel activity.

This cross-channel activity pattern is effectively illustrated by the vibrotactile waveforms at the end of the CVC utterance /s i s/ in Figure 4-6 (top left). The unvoiced final /s/ is characterized by a low-frequency, high-amplitude component and an aperiodic, high-frequency, low-amplitude component on the I-channel, similar to that observed at the beginning of the syllable, but the characteristic I-channel activity terminates more than 300 ms beyond the offset of voicing. The vibrotactile patterns for the syllable-final /t/, /f/, and /tʃ/ exhibit a similar relationship to their syllable-initial counterparts.

By contrast, the terminal /k/ and /p/ in Figure 4-6 (bottom left and top second from right, respectively) each exhibit a distinct, multi-channel offset pattern. In particular, the signals for both consonants include a terminal high-frequency component on the M-channel, reflecting spectral content in the 50-400 Hz

acoustic band. For both /k/ and /p/, the terminal M-channel (and T-channel) signals are in fact aperiodic, reflecting low-frequency acoustic noise, and they evoke a percept quite distinct from the periodic vibration of voicing. Such M-channel and T-channel components do not produce precisely localized percepts at the middle finger and thumb. They do, however, combine with the I-channel component to produce a tactual percept that is distinctly different from that produced by the I-channel component in isolation. The net result is a fairly robust vibrotactile cue for the articulatory place of these final consonants. Even in the absence of visual speechreading input, the tactile percepts of syllable-final /k/ and /p/ are distinct from that of syllable-final /d/. Since the low-frequency I-channel component for terminal /p/ is typically smaller than that for /k/, the glottal and labial final consonants are also tactually distinct from one another.

Voiced consonants in syllable-final position are distinguished from their unvoiced counterparts by the relatively concurrent offset of activity on the I-channel and periodic (voicing F₀) vibration on the M-channel. Depending on phonetic context and the degree of enunciation, the terminal voiced stop consonants and affricate may exhibit a pronounced lull in activity across channels, corresponding to the closure (or partial closure) preceding the consonant burst. From inspection of the labial and glottal plosive waveforms in Figure 4-6, one might infer that a pronounced lull on the M-channel might cause confusion between /b/ and /p/, or between /g/ and /k/, in syllable-final position. However, the post-release portions of terminal /b/ and /g/ are readily identified as belonging to voiced consonants due to the distinctive tactual percept of periodic vibration.

It is of interest to note that, using the transduction scheme of Yuan et al. (2004b, 2005b), the M-channel activity at the syllable-final /k/ and /p/ offsets does not provide a comparable articulatory place cue. On the contrary, the spectral characteristics of these terminal consonants have a particularly confounding influence under that scheme. Yuan et al. extracted energy (amplitude) envelopes of low-frequency (<350 Hz) and high-frequency (>3000 Hz) acoustic bands, which they used to modulate the amplitudes of a fixed frequency vibratory carrier signals. Their approach does not distinguish periodic from aperiodic acoustic signals, but conveys only the energy in each acoustic band. Thus, the low-frequency acoustic noise components of syllable-final /k/ and /p/ might easily be confused with the low-frequency acoustic voicing components of syllable-final /g/ and /b/, rendering these two consonant voicing pairs quite difficult to discriminate.

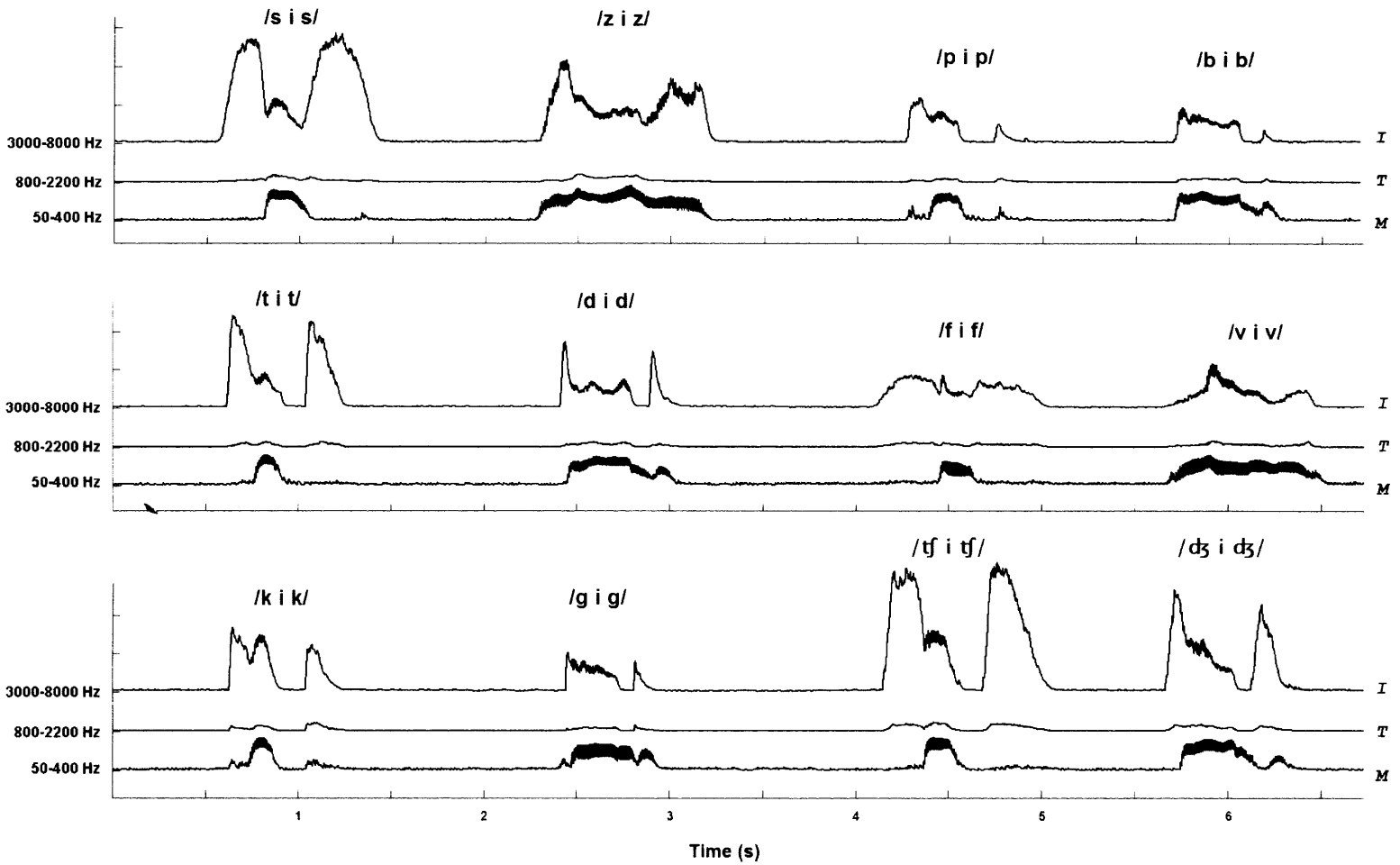


Figure 4-6. Three-channel vibrotactile signals corresponding to 12 CVC utterances produced by an adult male, each representing one of the 12 consonants included in the present study. Each CVC includes one consonant in both initial and final position, separated by the vowel /i/. Consonant voicing contrast pairs are positioned adjacent to one another. The letters *I*, *T*, and *M* to the right of each waveform reference the index finger, thumb, and middle finger, respectively, to which the signals were delivered.

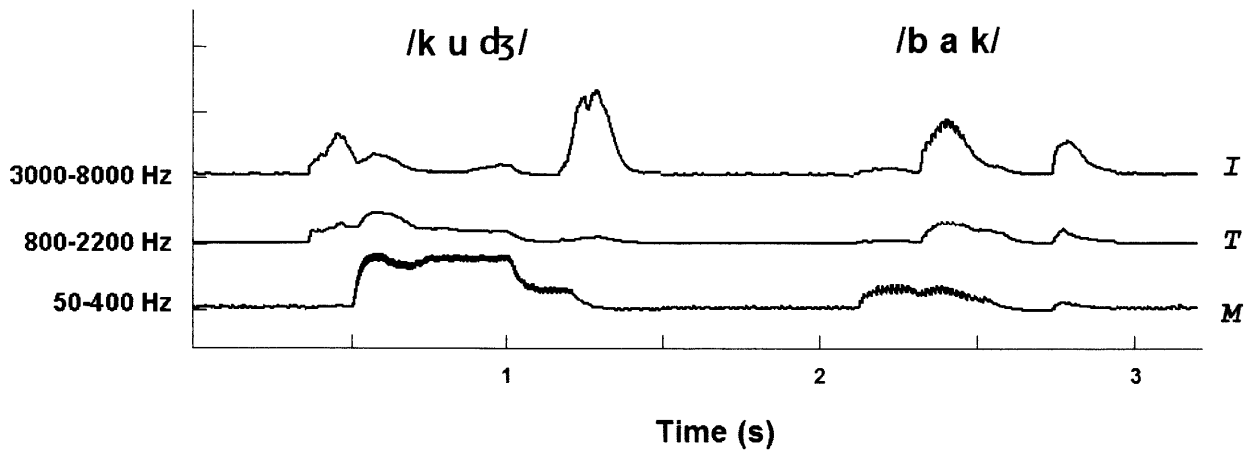


Figure 4-7. Three-channel vibrotactile signals corresponding to two recorded CVC utterances, spoken by the same adult female speaker. Note pre-voicing of initial /b/ evident on M-channel.

When the vibrotactile cues described in this section are combined with visual cues available through speechreading, it becomes possible to distinguish among the syllable-final consonants of Figure 4-6 nearly perfectly. The remarkably complementary interaction of visual and tactual cues that can so effectively enable final consonant identification is discussed further in Chapter 5.

4.6 Speech Processing Software

Audio signal processing software was written in the C++ programming language to perform the real-time processing of speech into vibrotactile control signals as illustrated in Figure 4-3. This software replaces a dedicated DSP hardware configuration described by Yuan (2003). The new software-based system offers more flexible online control of a wide range of basic operating and signal processing parameters, which has proven conducive to the development and optimization of tactual speech encoding. In particular, cutoff frequencies and gains associated with each filter indicated in Figure 4-3 can be varied independently (or the filter can be switched on/off) while the software is running, allowing for online adjustments and assessments that were not previously possible. The current system also offers increased compatibility with source code and other programming resources available in the public domain, which has benefited implementation and should continue to facilitate long-term development. With minor code modifications, the present DSP system can be implemented on most personal computers and some mobile

computing platforms, to process real-time microphone input (or recorded audio materials) into the vibrotactile control signals described herein.

Key elements of the speech processing system currently include the following:

- (1) PortAudio is a free, open-source, cross-platform, audio input/output library, written in the C programming language (<http://www.portaudio.com/>). The PortAudio project aims to support the development of audio software that is compatible with multiple computer operating systems.
- (2) The Audio Stream Input/Output (ASIO) soundcard driver protocol allows an application running on certain Microsoft Windows operating systems to bypass the standard DirectSound audio interface. Due largely to its reliance on large, fixed-size audio buffers, the DirectSound interface introduces long latencies to audio input and output processes, commonly resulting in processing delays of several hundred milliseconds. By contrast, communicating with the computer sound card via ASIO allows for input and output latencies on the order of 5-10 ms, thus making software-based processing in the Windows operating environment practicable for real time audio applications, which previously required the use of external, dedicated processing hardware. The ASIO application programming interface (API) is included as part of the ASIO software development kit (SDK) distributed by Steinberg Media Technologies (<http://www.steinberg.net>). Although available free of charge, the ASIO SDK is not open-source software, and certain licensing conditions apply to its usage. Note that inclusion of ASIO is required only for implementation on machines running Microsoft Windows --- the audio interfaces native to Linux distribution and Mac OS X operating systems, for example, do not produce comparable input and output latencies comparable to DirectSound.
- (3) The M-Audio Delta 1010LT multi-channel PCI soundcard offers eight analog inputs and eight analog outputs, as well as manufacturer-provided low-latency driver support. Until recently, use of the ASIO low-latency input/output protocol commonly required an ASIO soundcard driver tailored specifically to one's audio hardware. However, universal ASIO drivers that support most commercial soundcards and integrated audio processors are now freely available (e.g., <http://www.asio4all.com>). Of the available Delta 1010LT I/O channels, three analog outputs (controlling the three tactual stimulators) and a single analog input channel are used by the current system, leaving the remaining channels available for future development.

Chapter 5.

Tactile Speech Training and Evaluation

5.1 Adapting the Tadoma Lesson Plan

When deafblind children were instructed in the Tadoma method of tactual speechreading, perceptual training was closely related to speech production. The child's hand would move frequently between the face of the teacher and his/her own face, and the child was trained to produce speech by way of imitation, through thousands of repetitions (Stenquist, 1974). However, studies of artificial tactile speech strategies have focused on a subject's ability to discriminate a variety of tactually-encoded speech (or speech-like) stimuli, while the subject's own speech has been neglected.

Several factors underlie the tendency to avoid subject-produced speech stimuli. First, it is considered desirable to use identical stimuli with multiple subjects in a given experiment. Deaf speakers often have difficulty coordinating the production of certain phonemes (Hudgins, 1934; Lane and Perkell, 2005), and as such, speech sounds produced by deaf individuals may not be considered optimal training targets. Also, since many experiments are carried out with hearing subjects, strict measures are generally taken to ensure that tactile stimuli are not audible. Perhaps due to such factors as these, tactile speech researchers have not sought to model natural speech acquisition, such as commonly occurs in normal-hearing infants, in which vocal production and speech acquisition clearly go hand-in-hand.

Nearly a century after the first recorded efforts to develop artificial tactile speech communication for the deaf (e.g., Gault, 1924, 1926), even the most effective among a diversity of technological approaches has fallen well short of the success achieved by deafblind Tadoma users (Reed, 1995). To the extent possible, it would now seem reasonable to follow the lesson plans laid out by those teachers of the deaf and deafblind who pioneered the instruction of tactual speechreading during the first half of the 20th century.

Sophia K. Alcorn is the individual most commonly credited with developing the theoretical and educational principles of the Tadoma method, particularly in work carried out with her first two deafblind students, Tad Chapman and Oma Simpson, after whom Alcorn named the Tadoma method (Stenquist,

1974). Alcorn herself indicated that her work with Oma and Tad was grounded in lectures given by Caroline A. Yale at the Clarke School for the Deaf in Northampton, MA (Enerstvedt, 2000). Yale's publication of the "Northampton Vowel and Consonant Charts" and teachings concerning the instruction of deaf pupils in the development of elementary sounds of English (Yale, 1892; reprinted 1946) held fundamental importance to the early practitioners of Tadoma, and to many others in the area of deaf education. In addition to the writings of Alcorn (1932) and Yale, the written accounts of two other teachers, Rose M. Vivian (1966) and Gertrude Stenquist (1974), are particularly informative and useful in devising an approach to the training of artificial tactual speech. Stenquist's book provides a strong social context for the acquisition and effective use of tactual speech, and also quotes and describes materials from "An Outline of Work for the Deaf-Blind", prepared by Inis B. Hall, the first director of the Deafblind Department at Perkins School for the Blind (Watertown, MA) as an internal instructive document.

These instructor-caregivers spent much of their careers teaching deaf and deafblind children and adolescents to perceive speech as patterns of movement and vibration on the skin. The repeatability of their methods is beyond question, and their achievements have yet to be matched in conjunction with any artificial tactual display. Although published accounts of Tadoma teaching methods are somewhat sparse, the descriptions and recommendations left by these women are highly consistent with one another and certainly provide the basis for a systematic approach to training tactual speech reception.

A scientifically-grounded understanding of the efficacy of Tadoma training methods can be gleaned from the motor theories of speech perception, which are reviewed in Chapter 1 of this dissertation. Without subscribing to each and every dogmatic claim associated with the "Motor Theory", one may recognize the insightfulness of various arguments and interpretations put forward by Liberman and others. Thus, one objective of the training methods described below is to evaluate the hypothesis that the artificial introduction of speech through the somatosensory modality will benefit from the conditional association of patterned tactual inputs with corresponding articulatory gestures and proprioceptive activation patterns. During articulation, tactual speech users should thereby come to expect strong correlations between specific motor intentions, patterns of natural proprioceptive feedback, and artificial tactual feedback. Their familiarity with the patterns of speech-related tactile stimulation should allow fluent (if imperfect) imitation of a tactile voice signal.

We therefore sought to implement a sensorimotor tactual speech training strategy that pragmatically incorporates the subject's own speech into the experimental protocol. In addition to pressing buttons to

indicate judgments, subjects engaging in sensorimotor-enhanced training responded vocally and their vocalizations were transduced tactually. They were thereby provided frequent opportunities to compare the tactile percepts associated with the experimental speech stimuli and those deriving from their own vocalizations. Thus, under the sensorimotor training strategy, tactual cues were incorporated into the speech experience by eliciting vocal imitation of those cues, with the goal of enriching and facilitating perceptual learning of speech cues through the association of novel tactual sensations with the corresponding articulatory acts.

5.2 Methods

In all experiments, subjects sat facing a computer screen, which provided visual cues and feedback as appropriate. During experimental runs that included tactile stimulation, the subject's left hand rested on the tactual apparatus (described above), with the thumb, index, and middle fingers contacting the stimulators. During all experimental runs, the subject's right hand operated a computer mouse, whereby button-click responses were entered. During testing, profoundly deaf subjects removed any sensory aids normally worn. Both normal-hearing and profoundly deaf subjects wore foam earplugs (approximately 30 dB attenuating) and headphones delivering pink masking noise (roughly 80 dB SPL) in order to minimize the possibility of sounds from the device influencing performance. During sessions where the headset microphone was not needed, the subject with bilateral cochlear implants was not required to wear headphones.

5.2.1 Subjects

As evident in Table 5-1, experimental subjects included five profoundly deaf adults and three normal-hearing adults, ranging in age from 21-55 years. Profoundly deaf subjects included five females, all of whom reported spoken English as their primary mode of early communication. Normal-hearing subjects included one female and two males, all native speakers of American English (one male was raised speaking Arabic as well). Note that subject NH3 is the author. Each participant either provided a copy of a recent audiogram or underwent routine audiological testing in our laboratory prior to participating in any tactual experiments.

Table 5-1. Normal-hearing and profoundly deaf experimental subjects.

Subject	Age	Sex	Hearing Status	HA/CI (current)	Early Communication	Etiology
NH1	34	F	normal	--	spoken English	--
NH2	21	M	normal	--	spoken English/Arabic	--
NH3	34	M	normal	--	spoken English	--
HI1	32	F	profoundly deaf	hearing aids (bilateral)	spoken English	late onset / unknown (possibly autoimmune)
HI2	31	F	profoundly deaf	bilateral CI (first at age 19)	spoken English (w/ amplification)	meningitis (age 2), then progressive
HI3	30	F	profoundly deaf	CI (since age 16)	spoken English (w/ amplification)	congenital
HI4	55	F	profoundly deaf	CI (since age 49)	spoken English (w/ amplification)	congenital
HI5	53	F	profoundly deaf	none	spoken English (w/ amplification)	congenital / maternal Rubella

5.2.2 Tactual Psychophysics

Tactual Detection

To determine subjects' thresholds for detection of tactual stimulation, measurements were taken individually for the thumb, index finger, and middle finger of the left hand. For each measurement, sinusoidal vibratory stimuli were delivered to the distal glabrous surface of the digit. Stimuli had 25-ms rise/fall times and were 500 ms in duration, sufficiently long such that detection threshold level should be independent of duration (Verrillo, 1965). The frequencies examined were 10, 50, and 200 Hz, spanning most of the human tactual perceptual range for vibratory stimuli that is engaged by the tactile speech display under evaluation.

Detection thresholds were measured using a two-interval, two-alternative forced choice (2I-2AFC) adaptive procedure. Subjects were asked to indicate in which of two visually cued intervals a tactual stimulus was presented. They received visual correct-answer feedback for each trial. Following an initial supra-threshold stimulus, subsequent stimulus presentations were governed by a "two-down, one-up" paradigm, which is expected to converge upon the stimulus level at which a subject responds correctly

70.7% of the time (Levitt, 1971). An initial amplitude step-size of 5 dB was decreased to 2 dB following the second reversal in the direction of stimulus amplitude adjustment. Following the third reversal, the step-size was decreased to 1 dB and held this value for the remainder of the run. Each run was terminated following the tenth reversal of direction, and the detection threshold was defined as the average stimulus level over the final six reversals. At each digit, and for each vibratory frequency, this protocol was repeated at least twice, and the results of the two runs were averaged. Up to three additional measurements were taken in cases where the initial two measurements differed by more than 5-6 dB or deviated by more than 10 dB from the mean thresholds recorded in Chapter 3 of this dissertation.

Tactual Onset-order Discrimination

Temporal onset-order discrimination was examined using a one-interval, two-alternative forced choice (1I-2AFC) adaptive procedure. During each trial, 200 Hz sinusoidal vibrations with 10-ms rise/fall times were presented at a level of 15 dB re 1 μ m peak displacement to the distal glabrous surfaces of both the middle and index fingers of the subject's left hand. In each case, the onset of stimulation to one finger preceded the onset of stimulation to the other finger, such that the stimulus onset asynchrony (SOA) can be defined as the difference in stimulus onset timings, [$\text{Onset}_{\text{Index}} - \text{Onset}_{\text{Middle}}$]. Thus, a negative SOA value ($\text{SOA} < 0$) in a given trial indicates that the index finger stimulus onset preceded that of the middle finger, whereas a positive SOA value ($\text{SOA} > 0$) indicates that the middle finger onset preceded that of the index finger. Stimulation of the two fingers always terminated simultaneously, 800 ms after the earlier onset time (i.e., the longer of the two stimuli had a total duration of 800 ms).

Following each presentation, subjects were asked to indicate which stimulus had the earlier onset. They were instructed to use a computer mouse to select a graphic button labeled "Middle" if the onset of the stimulus delivered to the middle preceded that of the stimulus delivered to the index finger, and to select the button labeled "Index" if the index finger stimulus onset arrived earlier. Subjects received visual trial-by-trial correct-answer feedback.

On any given trial, the sign of the SOA (i.e., the identity of the finger receiving the earlier-onset stimulus) was randomly determined, and the absolute value of the SOA ($|\text{SOA}|$) was determined according to a "two-down, one-up" adaptive paradigm. The initial $|\text{SOA}|$ was set at either 200, 250 or 300 ms, the lowest of these values that proved sufficiently large for the subject to readily distinguish the identity of the earlier-onset stimulus. An initial $|\text{SOA}|$ step-size of 20 ms was decreased to 10 ms following the first reversal in the direction of $|\text{SOA}|$ duration adjustment. The step-size was decreased to 5 ms following the

fourth reversal and to 2 ms following the seventh reversal, holding the latter value for the remainder of the run. Each run was terminated following the fifteenth reversal of direction, and average |SOA| over the final six reversals was taken as the single-run threshold measurement. Each subject repeated this protocol no fewer than 12 (and no more than 20) times, and a subject's onset-order discrimination threshold was calculated as the mean of these measurements.

5.2.3 Speech Stimuli

Speech stimuli consisted of brief video recordings of two adult female speakers producing isolated CVC nonsense syllables (frontal view of head/neck against solid backdrop). Vowels were restricted to the three English cardinal vowels /a/, /i/, /u/, and consonants in both syllable-initial and syllable final positions included the 12 English consonants /b/, /p/, /d/, /t/, /g/, /k/, /v/, /f/, /z/, /s/, /dʒ/, /tʃ/. A total of 844 distinct CVC videos, stored as MOV (Apple QuickTime format) files were used. Of these, approximately two-thirds served as "training stimuli" and were presented along with correct answer feedback during the various experimental phases. The remaining one-third of CVC videos were reserved as "evaluation stimuli", which were only presented without feedback.

5.2.4 Training and Evaluation Protocol

The last three rows of Table 5-2 outline the three primary phases of the tactual speech training and evaluation protocol. Prior to the "pairwise discrimination" phase, subjects performed a rudimentary vowel identification task, which served to familiarize them with the experimental apparatus and the perceptual diversity they might expect from tactual speech stimuli in this study. The three subsequent phases entailed systematic training of consonant distinctions, progressing from individual pairwise consonant discriminations to complete CVC identification. Formal psychophysical evaluations of acquired perceptual skills were performed upon completion of training during the "pairwise discrimination" and "CVC identification" phases only.

Training Strategies

All training trials began with two presentations of a given CVC stimulus. The first presentation included both video and tactual components (V+T condition). The audio was transduced as described in Chapter 4 of this document. The second presentation included only the tactual component (T-alone, not accompanied by video image). In all training trials, subjects' mouse-click responses were followed immediately by correct-answer feedback. The graphic button corresponding to the correct answer was highlighted in bright red when the subject responded incorrectly, and it was highlighted in bright green

Table 5-2. Outline of tactual speech experimental phases.

Phase	Stimulus groupings	Objective	Experimental Procedure
Familiarization	CVC groupings differing in vowel only	Introduction to apparatus / stimulus variability	informal
Pairwise Discrimination	/ t - s / / z - s / / t - d / / p - b / / k - g / / f - v / / tʃ - dʒ /	TRAINING and EVALUATION	1-interval 2-AFC CI and CF separately
Six-consonant Groupings	/p b f v tʃ dʒ/ /k g t d s z/	TRAINING ONLY	1-interval 6-AFC CI and CF separately
CVC Identification	CVC-Initial & CVC-Final /p b t d k g f v s z tʃ dʒ/	TRAINING and EVALUATION	1-interval 12,12-AFC CI and CF <u>together</u>

when the subject responded correctly. In trials where the subject responded incorrectly, the correct answer remained highlighted in red as the tactual component of the CVC stimulus was repeated once more (T-alone condition).

In the second through fourth experimental phases, consonant discrimination and identification training were performed according to one of two training strategies. Under the *non-vocal* (NV) strategy, subjects trained without vocalizing for tactual feedback, providing only button-responses for each trial, as required. Under the *vocal-feedback* (VF) strategy, each trial additionally included a "vocal response period" following the stimulus presentation but preceding the button-response. During the vocal response period, subjects produced monosyllabic utterances, intended to approximate the CVC stimulus and to vocally explore key consonant feature contrasts. Vocalizations were picked up by a headset microphone and transduced for tactual display in real time. Subjects were encouraged to compare their own utterances with the CVC stimulus, thereby using vocal feedback to inform perceptual judgments.

Subjects were alternately designated to complete training according to either the VF or NV strategy, based on the order in which they entered the study. After completing the full experimental protocol according to the assigned training protocol, most subjects then repeated the consonant discrimination and identification phases a second time, according to the other training protocol.

Note that regardless of the training strategy employed, formal evaluations always used a simple one-interval, N alternative forced choice procedure. During evaluation trials, the CVC stimulus was presented only once, subjects did not vocalize, and correct-answer feedback was not provided.

Phase 1: "Familiarization" through Vowel Identification

"Keeping continually in mind the learning process of the little hearing baby, we do not allow the deaf child to attempt voice until he is thoroughly saturated with the way the voice of the teacher feels..."

-- Sophia Alcorn, Tadoma speechreading pioneer (1932)

A primary objective of the *familiarization* phase is that subjects become "thoroughly saturated" with the feel of the tactual apparatus and the range of vibrotactile outputs. The vowels /u/, /a/, and /i/ are easily distinguished by naive and experienced subjects alike when the speaker's face is visible. These three vowels also have distinct tactual representations (depicted in Figure 4-5). Familiarity with each vowel's unique tactual percept was intended to enable subjects to discern the tactual correlates of coarticulated consonants more effectively (i.e., to appreciate each consonant as an operator on its vowel environment, rather than attending strictly to invariant characteristics).

The distinct tactual features of each of the three vowels were first introduced to subjects through manual presentations of sets of three CVC stimuli that differed only by their center vowel (i.e., all CVCs in a set consisted of the same initial and final consonants, bounding one of the vowels /u/, /a/, and /i/). These presentations included only the transduced tactile component of the CVC training video clips (i.e., T-alone condition). Three or four such sets were presented, including a variety of voiced and unvoiced consonants, such that the subject could distinguish the "steady-state" portion of each vowel. Subjects were permitted to play each set of CVCs several times for comparison.

Subjects then performed a one-interval, three alternative forced-choice (1I-3AFC) vowel identification task. Stimuli were chosen randomly from among the training set of CVCs, including both voiced and unvoiced consonants in combination with the vowels /u/, /a/, and /i/. Each subject performed one experimental run of 40 trials of under the *visual+tactile* (V+T) condition, during which vowels were easily distinguishable via visual cues. Next, they performed the task in the *tactile-alone* (T) condition, completing at least four runs of 40 trials each. Under both conditions, subjects received trial-by-trial correct-answer feedback.

Phase 2: Pairwise Consonant Discrimination

As indicated in Table 5-2, subjects were trained to discriminate seven consonant pairs, including six voicing contrasts (/p-b/, /t-d/, /k-g/, /f-v/, /s-z/, /tʃ-dʒ/) and one manner contrast (/t-s/). Training was performed according to a one-interval 2AFC protocol, modified as described above. Discrimination of each consonant pair was trained separately in the syllable-initial and syllable-final positions. Generally, for each pair, in each position, subjects performed one 40-trial run under the *visual-alone* (V-alone) condition, followed by four 40-trial runs under the V+T condition. After completing one cycle through the discrimination pairs, subjects cycled through them again, completing up to four additional 40-trial runs under the V+T condition. For any given consonant pair/position, if a subject scored 35 or more correct in consecutive 40-trial V+T runs, no further V+T training runs were conducted.

For training under the *vocal feedback* strategy, the standard mouse-click response in each trial was preceded by a "vocal response period", during which subjects were instructed to attempt at least one vocal imitation of the CVC stimulus. Subjects' vocalizations were picked up by a microphone and processed via the same DSP transduction system as the audio portion of the CVC stimuli, such that they received tactual feedback of all vocalizations in real time. Subjects were asked to attend closely to tactual feedback of their own vocalizations, comparing them with the tactual percept accompanying the CVC stimulus presentation. The vocalization might be a partial reproduction of the perceived CVC, including only the CV or VC portion of the syllable (i.e., consisting of the perceived vowel combined with the putative target consonant). Alternatively, the vocalization might be a complete CVC, including the non-target consonant as well as the target consonant. If at all uncertain about the identity of the target consonant, subjects were encouraged to produce alternative vocalizations for comparison, one with each member of the discrimination pair. No upper limit was placed on the number of vocalizations produced or the duration of the vocal response period. Subjects had the option of repeating a syllable several times or alternating between a voiced and an unvoiced consonant. The only explicit requirement during the vocal response period was the production of at least one vocalized syllable. Anytime thereafter, subjects could terminate the vocal period by clicking on one of the two graphic response-buttons to indicate the identity of the target consonant.

Subjects received correct-answer feedback immediately following the mouse-click response. In trials where they responded incorrectly, the correct-answer indicator remained on-screen as the tactual component of the CVC stimulus was presented once more. The beginning of a new trial was indicated by the disappearance of the feedback indicator, followed by the presentation of a new CVC stimulus. At the end of each run, the number of correct responses and the number of trials total provided subjects with

cumulative feedback for the preceding run. Each new score was entered onto a score sheet, which subjects could view at any time. Each experimental run also produced a detailed record of trial-by-trial performance data, which was saved as a computer text file for offline analysis.

Upon completion of training, formal evaluation was performed using a one-interval 2AFC (1I-2AFC) protocol without trial-by-trial feedback, using a separate but comparably composed set of CVC stimuli (the "CVC Evaluation set"). During the evaluation period, subjects did not respond vocally or attempt to imitate the stimuli, regardless of the strategy by which they had trained. A single 80-trial evaluation run was performed for each consonant pair/position, under each of the V-alone and V+T conditions.

For purposes of analysis, data for each consonant pair/position combination under each experimental condition were summarized (offline) in a 2x2 stimulus-response confusion matrix, and the signal detection measures of sensitivity (d') and bias (β) were calculated (Green and Swets, 1966; Durlach, 1968; Macmillan, 1990). For a given block of 1I-2AFC trials, the confusion matrix served to tabulate the number of trials corresponding to each possible combination of stimuli (S1 or S2) and responses (R1 or R2). From this matrix, a "hit rate" (H) and a "false-alarm rate" (F) were calculated as

$$H = S1R1 / (S1R1 + S1R2) \quad F = S2R1 / (S2R1 + S2R2)$$

From these values, d' and β were then calculated according to the formulae

$$d' = z(H) - z(F) \quad \beta = \frac{z(H) + z(F)}{2}$$

where the function $z(x)$ is the inverse of the normal distribution function, which converts x to a z-score.

Phase 3: Six-consonant Groupings

For the third experimental phase, the 12 consonants of the *CVC Training Set* were separated into two groups of six: /p b f v tʃ dʒ/ and /k g t d s z/. During any given run, target consonants were restricted to one of these two groupings and to either syllable-initial or syllable-final position. Subjects trained to identify targets from among the six possible alternatives. As indicated in Table 5-2, the *Six-consonant Groupings* phase consisted entirely of training trials, performed under the tactually supplemented lipreading (V+T) condition. The specific groupings of consonants were chosen to introduce new

articulatory feature contrasts (e.g., /d-z/ and /d-g/) in a context that builds upon and reinforces distinctions learned during the pairwise discrimination phase, particularly the more perceptually challenging of the voicing contrasts (e.g., /tʃ-dʒ/ and /f-v/).

Subjects performed up to 400 V+T training trials for each of the four possible combinations of consonant grouping and position, divided into runs of 50 trials each. In general, four 50-trial runs of each type were performed in succession, before alternating groupings. For any given consonant grouping, if a subject scored 44 or more correct in consecutive 50-trial V+T training runs, no further runs of that type were conducted. For subjects training under the *vocal feedback* strategy, all Phase 3 trials included a vocal feedback component, following the same vocal and manual response procedure described for the previous phase.

Phase 4: CVC Identification

During the *CVC Identification* phase, subjects trained to identify both consonants of each CVC stimulus in the course of a single trial. In any given trial, both the initial and final consonants of the target CVC stimulus varied among the 12 consonants (/p b t d k g f v s z tʃ dʒ/) encountered in prior training, with the central vowel restricted to /u/, /a/, and /i/. The task was effectively a compounding of two single-interval 12-alternative forced choice procedures. Following presentation of the CVC stimulus, subjects provided two 12-alternative responses in succession, one to identify the initial consonant and the other to identify the final consonant. The two responses were entered in two spatially distinct graphical response button grids. Each grid included 12 clearly marked buttons, one for each consonant choice. The response-button configuration was kept constant throughout the experiment.

Training was conducted in runs of 50 trials each, in which correct-answer feedback was provided immediately following each button-response. If either the initial or final consonant was incorrectly identified, the T-alone stimulus was repeated once more before the trial was terminated. For subjects training under the *vocal feedback* strategy, the vocal component of CVC identification trials followed essentially the same procedure described for the previous phases. Training included V-alone and V+T trials in the ratio of approximately 1:5, for a maximum of 1200 trials total.

Formal evaluation was performed in 50-trial runs, alternating between the V and V+T conditions. Most subjects performed 15 evaluation runs under each of the V-alone and V+T conditions. Evaluation trials included a single presentation of a CVC stimulus chosen at random from the evaluation set, following

which subjects provided two button-click responses in succession, without feedback. Subjects did not vocalize during evaluation runs, regardless of their assigned training condition.

For purposes of data analysis, pooled results for each consonant position and experimental condition were summarized (offline) in a 16x16 stimulus-response confusion matrix and analyzed in terms of percent correct and information transfer (IT). Information transfer provides a measure of the covariance between the stimuli and responses (i.e., the accuracy with which the stimulus predicts the response). Moreover, by grouping stimuli and responses by articulatory feature (i.e., voicing, place, or manner), condensed confusion matrices were constructed, and percent correct and IT scores were calculated for each feature individually (Miller and Nicely, 1955), thereby permitting comparison of subjects' reception of each individual articulatory feature under the V and V+T conditions.

5.3 Results

5.3.1 Tactual Sensitivity

Figure 5-1 shows tactual detection thresholds measured at each subject's left thumb, index, and middle fingers, using 500-ms sinusoidal vibratory pulses at frequencies of 10 Hz, 50 Hz, and 200 Hz. Asterisk markers indicate the average threshold across subjects for each stimulation frequency, measured at the distal glabrous surface of each digit, and error bars indicate one standard deviation. Mean and standard deviation values for tactual detection thresholds at each site, for each frequency, are also listed in Table 5-3. Note that no data are shown for subject HI3, who was not available for tactual sensitivity or temporal resolution measurements (performed subsequent to tactile speech training).

Detection thresholds for 10 Hz sinusoidal vibrations, shown in the leftmost plot of Figure 5-1, fall in the range of 19 to 36 dB re 1 μ m peak displacement, with mean thresholds ranging from 25.2 dB to 26.9 dB for the three fingers. Most individual measurements lie within 5 dB of the mean range, although measurements for subjects NH2 and HI4 are notable exceptions. At the index finger, subject HI4 has a detection threshold of 36.0 dB re 1 μ m peak, which is substantially elevated relative to the group mean of 26.8 dB (\pm 5.2 dB s.d.) and suggests reduced sensitivity to 10 Hz vibration. By contrast, thresholds for subject NH2 at the left index finger and thumb (19.1 dB and 19.5 dB, respectively) suggest heightened sensitivity to 10 Hz vibration.

Detection thresholds for 50 Hz sinusoidal vibrations, shown in the center plot of Figure 5-1, range between 2.7 and 19.3 dB re 1 μm peak displacement at the low and high extremes. All measurements other than those two fall in the range of 6.3 to 14.3 dB re 1 μm peak displacement. Mean 50 Hz thresholds for the left thumb, middle, and index fingers are 9.7 dB, 8.6 dB, and 11.4 dB, respectively. The 2.7 dB threshold recorded for subject NH1 at the left thumb reflects the averaging of two measurements that differed substantially (7.7 dB and -2.3 dB), and so the reliability of this low threshold value is uncertain. The 19.3 dB threshold recorded for subject NH4 reflects the averaging of five separate measurements, and being nearly two standard deviations above the group mean, it suggests that NH4 exhibits reduced sensitivity to 50 Hz vibrations at the index finger.

Detection thresholds for 200 Hz sinusoidal vibrations, shown in the rightmost plot of Figure 5-1, range between -29.3 and -0.4 dB re 1 μm peak displacement at the low and high extremes. Mean 200 Hz thresholds for the left thumb, middle, and index fingers are -16.5 dB, -14.7 dB, and -16.8 dB, respectively. The -29.3 dB threshold recorded for subject NH3 at the left index finger reflects the averaging of two measurements (-27.5 and -31.2 dB). Thresholds recorded for subject HI2 at each finger reflects the averaging of four separate measurements. Note that all three 200 Hz thresholds for HI2 exceed their respective group means by more than one standard deviation. In particular, the -0.4 dB threshold recorded for the middle finger of HI2 exceeds the group mean by nearly two standard deviations. The -5.8 dB threshold recorded for the middle finger of subject HI4 reflects the averaging of five separate measurements, and thus also likely reflects reduced sensitivity to 200 Hz vibration. Note that the left index finger of subject HI4 exhibits threshold elevation for all three frequencies of sinusoidal vibration tested.

Table 5-3. Mean tactual threshold (and standard deviation) across subjects measured in dB re 1 μm peak displacement at the left thumb, index, and middle fingers, using sinusoidal vibratory frequencies of 10 Hz, 50 Hz, and 200 Hz.

Frequency	10 Hz threshold (dB re 1 μm)		50 Hz threshold (dB re 1 μm)		200 Hz threshold (dB re 1 μm)	
	mean	s.d.	mean	s.d.	mean	s.d.
Left Index Finger	26.8	5.2	11.4	4.3	-16.8	8.7
Left Middle Finger	26.9	3.5	8.6	2.5	-14.7	7.7
Left Thumb	25.2	4.1	9.7	3.5	-16.5	6.1

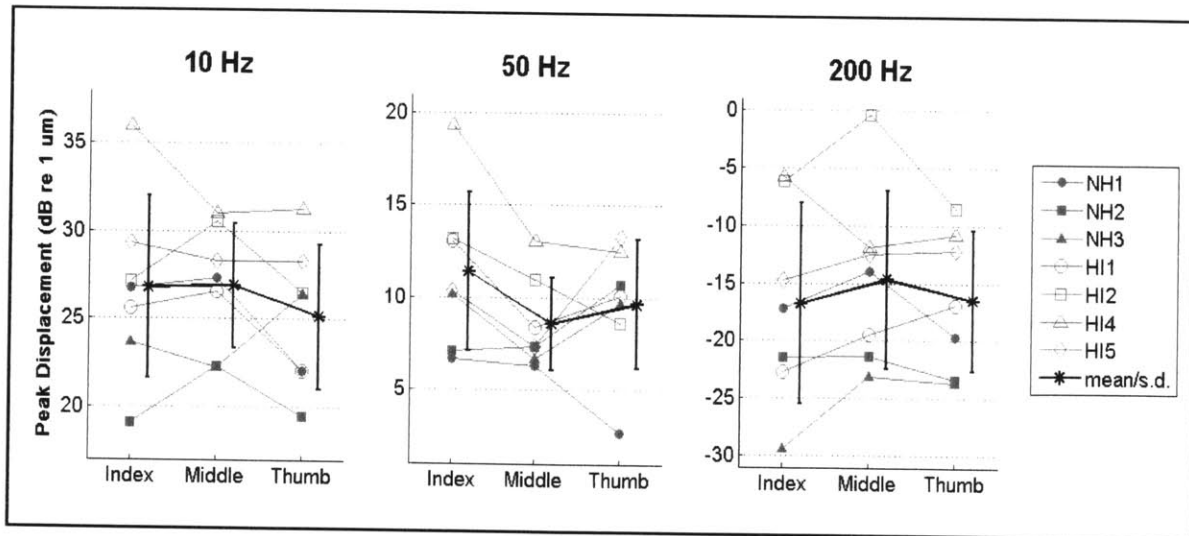


Figure 5-1. Tactual detection thresholds in dB re 1 μm peak displacement, measured at the distal glabrous surface of each subject's left index finger, middle finger, and thumb. Detection thresholds measured with sinusoidal vibratory stimuli at frequencies of 10 Hz, 50 Hz, and 200 Hz are shown in the left, middle, and right panels, respectively. Data points for each subject are distinguished by a unique marker symbol, as indicated in the figure legend. Asterisk symbols and associated error bars indicate group means and standard deviations for each finger-frequency combination.

5.3.2 Tactual Onset-Order Discrimination

Tactual onset-order discrimination was examined using an adaptive 1I-2AFC procedure, as described above in section 5.2.2. Briefly, overlapping 200 Hz sinusoidal vibrations were delivered to the left index and middle fingers, and subjects were asked to indicate which of two stimuli had an earlier onset.

Table 5-4 shows the temporal onset-order thresholds measured for seven of eight experimental subjects, along with the standard error of the mean (SEM). Each threshold value listed is the mean of between 12 and 20 measurements taken for the corresponding subject. SEM was calculated as the standard deviation of threshold measurements divided by the square root of the number of measurements taken for that subject.

The lowest observed onset-order threshold was 52.8 ms (subject NH3), and the highest observed was 294.1 ms (subject HI5). The average value of thresholds across subjects is 175.7 ms, with a standard deviation of 96.5 ms.

Table 5-4. Tactual onset-order discrimination thresholds (and SEM values) for seven out of eight subjects, measured using an adaptive 1I-2AFC procedure, converging on 70.7% correct response rate. (Subject HI3 was not available for testing.)

Subject	Onset-order discrimination Threshold (ms)	SEM (ms)
NH1	174.1	8.5
NH2	54.9	5.7
NH3	52.8	3.4
HI1	219.8	12.3
HI2	159.1	6.0
HI3	--	--
HI4	275.1	11.2
HI5	294.1	14.1
Mean ± s.d.	175.7 ± 96.5	

5.3.3 Pairwise Consonant Discrimination

During the pairwise consonant discrimination phase of each experimental repetition, subjects trained to discriminate seven consonant pairs, in both syllable-initial and syllable-final contexts, exploiting tactually-presented voicing and manner information. Discrimination performance was then evaluated for each consonant pair individually, under both visual-alone (V-alone) and visual+tactile (V+T) conditions.

Figure 5-2 presents mean d' values, averaged across subjects, reflecting the discriminability of each consonant training pair under V-alone and V+T conditions (dark and light bars, respectively). The seven leftmost labels on the abscissa correspond to discrimination of the specified consonants in syllable-initial context, while the seven labels to the right of the vertical dotted-line correspond to consonant

discrimination in syllable-final context. For both the top and bottom panels of Figure 5-2, data from the first (and only) pairwise evaluations for subjects NH3 and HI3 are used. However, for the six subjects who repeated the experimental procedure twice (using different training strategies), data from the first pairwise evaluation were included in the averages shown in top panel of Figure 5-2, while data from the second pairwise evaluation were included in the averages shown in the bottom panel.

For each consonant discrimination pair in the top and bottom panels of Figure 5-2, the set of d' values measured under the V-alone and V+T conditions were compared using a one-tailed T-test. Single (*) and double (**) asterisks indicate pairs for which the resulting p-value (probability of observed measurements if the V+T mean does not exceed V-alone mean) is less than 0.05 and 0.01, respectively.

Mean d' values shown in the bottom panel generally exceed the corresponding values in the top panel, but the overall performance pattern for each consonant pair under V-alone and V+T conditions is similar. The most notable exception is the improvement in tactual benefit for the consonant pair /f/ -/v/ in syllable-initial position (labeled FVi in Figure 5-2). Whereas, in the top panel of Figure 5-2, no statistical difference is observed between mean d' values for FVi in the V-alone and V+T conditions, the difference between the two conditions is significant (at $p < .01$) in the bottom panel, where the V+T mean score exceeds the discriminability criterion ($d' = 1$).

Two-tailed T-tests were performed to compare mean performance under like conditions (i.e., same consonant pair/position and sensory condition) in the subjects' first and final procedural repetitions. In no case was the difference between first and final repetition mean score found to be statistically significant at $p < .05$. These same T-tests were repeated, excluding the data of subjects NH3 and HI3 (for whom first repetitions were also their final repetitions), and again no significant differences were found.

Overall, Figure 5-2 reveals that, under the V-alone condition, syllable-initial consonant discrimination performance falls close to chance (i.e., $d' = 0$) for all seven consonant pairs. By contrast, in syllable-final position, near-chance performance is observed only for the /t/-/s/ consonant pair (labeled TSf), which is the only pair among the seven featuring an articulatory manner contrast (the other six consonant pairs are distinguished by the voicing feature).

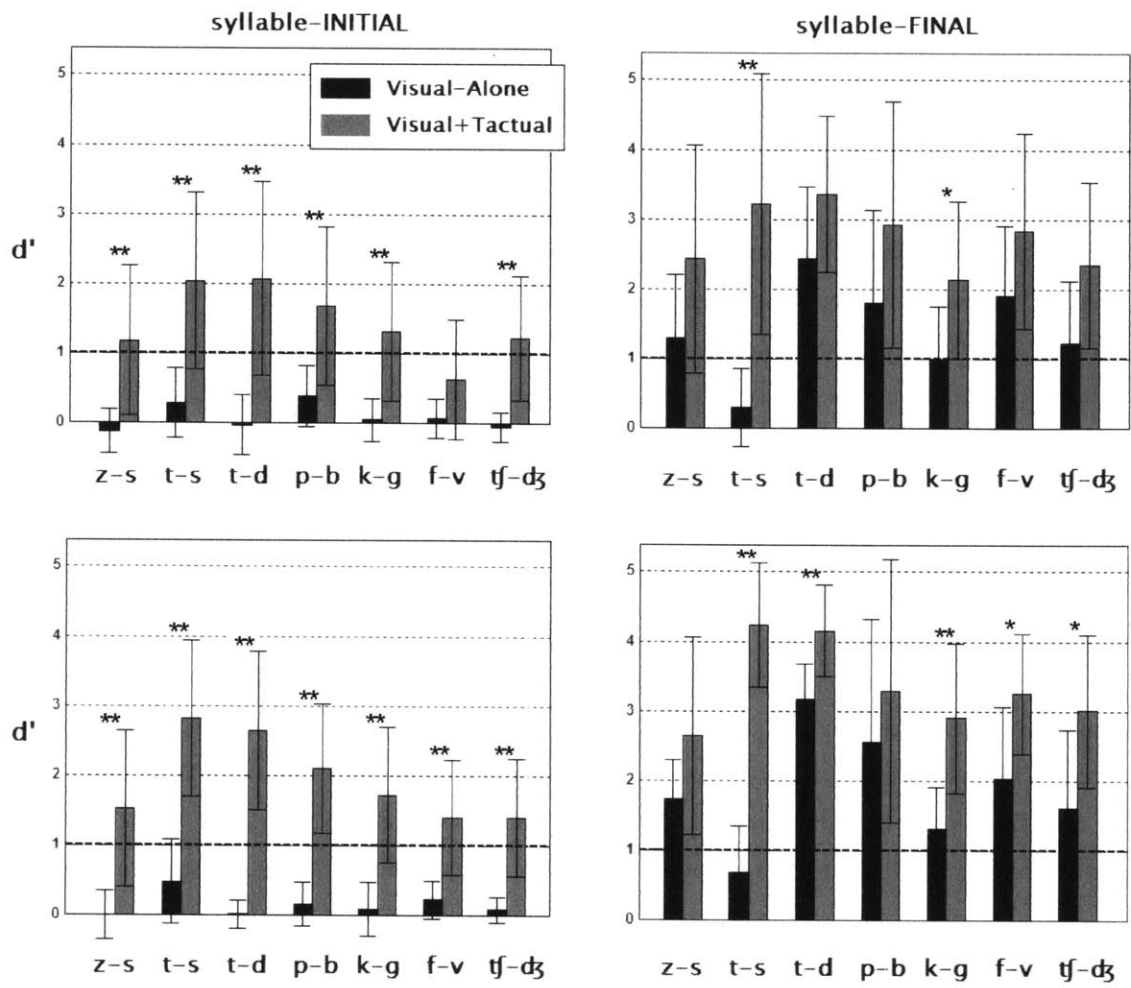
When visual and tactual information are provided together (V+T condition), mean subject performance improves substantially. In syllable-initial consonant discrimination, where all seven consonant pairs are discriminated poorly in the V-alone condition, the enhancement of performance under the V+T condition

is clearly evident. Performance exceeds the $d'=1$ criterion for six of seven consonant pairs in the top panel of Figure 5-2 and for all seven pairs in the bottom panel. In the latter instance, one-tailed T-test comparisons indicate that mean d' values for discrimination of each consonant pair under the V+T condition significantly exceed the corresponding V-alone values at a level of $p < .01$.

In syllable-final consonant discrimination, the most notable improvement in performance between the V-alone and V+T conditions (in both the top and bottom panels of Figure 5-2) is observed for the manner-contrast pair /t/-/s/, for which performance in the V condition is closest to chance. In the top panel, mean d' values for the V-alone and V+T conditions also differ significantly for the syllable-final voicing-contrast pair /k/-/g/ ($p < .05$), but not for any of the other five syllable-final voicing-contrast pairs, all of which are discriminated relatively well ($d' > 1$) in the V-alone condition. In the bottom panel of Figure 5-2, mean d' values for the V and V+T conditions differ significantly ($p < .05$) for all syllable-final consonant pairs other than /z/-/s/ and /p/-/b/.

Figure 5-3 presents the mean d' (top panel) and absolute bias (bottom panel) values, averaged across the seven initial consonant pairs, for each subject in each repetition of the pairwise consonant discrimination evaluation. Dark and light bars reflect measurements performed under the V-alone and V+T sensory conditions, respectively. Error bars indicate one standard deviation from the mean. "VF" and "NV" designations on abscissa indicate the type of training protocol (vocal-feedback or non-vocal) employed prior to evaluation in a given experimental repetition. Note that subjects NH3 and HI3 each performed one repetition only. For all subjects other than NH3 and HI3, results for two repetitions of the experiment are shown, one performed with the "vocal-feedback" training strategy (VF) and the other performed without vocalization (NV). For each subject label on the abscissa, the leftmost indicated training strategy (VF or NV) was employed during that subject's first repetition of the experimental protocol.

The top panel of Figure 5-3 shows that, under the V-alone sensory condition (dark bars), d' values fall close to zero for all subjects in all repetitions, reflecting subjects' inability to discriminate between the paired initial consonants via lipreading alone. Under the V+T condition (light bars), however, seven out of eight subjects achieve mean d' values exceeding the $d'=1$ criterion (indicated by dashed line). Six subjects in particular --- NH1, NH2, NH3, HI1, HI2, and HI3 --- exceed $d'=1$ in the V+T evaluation of their first procedural repetition, and of these, all but HI2 exceed $d'=2$ (roughly 85% correct response rate) in either the first or second repetition. Moreover, a one-tailed T-test reveals that, for all six of these subjects, in all repetitions, performance improvement under the V+T condition (relative to V-alone) is significant at a level of $p < .01$.



Consonant Discrimination Pairs

Figure 5-2. Mean d' values, averaged across all subjects, for each consonant discrimination pair in syllable-initial (leftmost seven pairs) and syllable-final positions (rightmost seven pairs). Top panel: data included derives from pairwise evaluation during each subject's first experimental repetition. Bottom panel: data included derives from each subject's final pairwise evaluation session (i.e., the second pairwise evaluation for subjects who performed two experimental repetitions). Error bars indicate one standard deviation from the mean. Dashed line at $d'=1$ corresponds to approximately 70% correct response rate. Dark and light bars reflect measurements performed under the V-alone and V+T conditions, respectively. Single (*) and double (**) asterisks indicate p-values ($p < .05$ and $p < .01$, respectively) for one-tailed T-tests comparing V-alone and V+T means for each consonant pair.

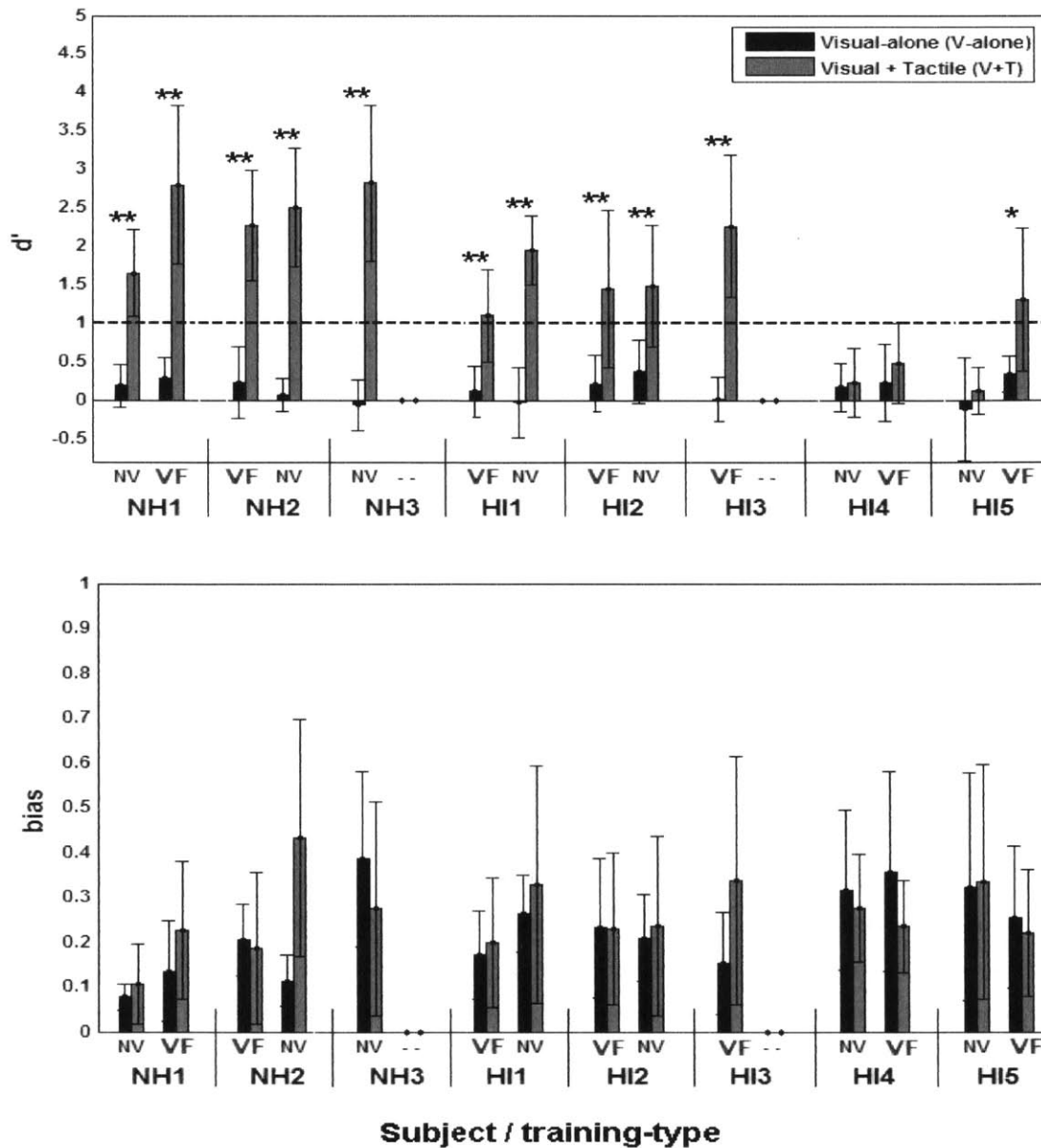
During her first repetition of the experiments, subject HI5 performs close to chance ($d'=0$) under both V-alone and V+T conditions in the pairwise initial consonant discrimination evaluation. However, HI5 exhibits much improved V+T performance during the second repetition, with her cross-pair average sensitivity surpassing the $d'=1$ criterion. In this second repetition, HI5's mean d' score under the V+T condition exceeds that under the V-alone condition at a significance level of $p<.05$ (as assessed by one-tailed T-test).

Of the eight subjects, only HI4 shows no notable performance enhancement under the V+T sensory condition in either her first or second procedural repetition. Mean d' scores for HI4 under the V-alone and V+T conditions do not differ significantly.

Figure 5-4 presents the mean d' (top panel) and mean absolute bias (bottom panel) values, averaged across the seven final consonant pairs, for each subject in each repetition of the pairwise consonant discrimination evaluation. Error bars indicate one standard deviation from the mean. Each set of dark and light bars correspond to discrimination performance of a given subject under the V-alone and V+T conditions, respectively, during one experimental repetition. The type of training received during that repetition (prior to evaluation) is indicated by the "VF" and "NV" designations on abscissa.

The top panel of Figure 5-4 reveals that, overall, pairwise discriminability of syllable-final consonants under the V-alone sensory condition (dark bars) is substantially better than that observed for syllable-initial consonants (in Figure 5-3). In fact, all subjects other than HI5 demonstrated mean sensitivity scores greater than $d'=1$ under the V-alone condition in their first repetitions, as well as in any subsequent repetitions performed. By contrast, subject HI5 exhibited chance performance under the V-alone condition during her first repetition of the pairwise discrimination experiment; her V-alone discrimination performance showed improvement in the second repetition, although her mean V-alone d' score still fell short of unity.

Under the V+T sensory condition, pairwise discriminability of syllable-final consonants improved substantially for most subjects. Subject HI5 performed substantially better under the V+T condition relative to V-alone, but only during her second repetition of the experiment; her first repetition V+T performance remained close to chance level. One-tailed T-tests comparing performance across consonant



NV = Non-vocal Training
 VF = Vocal-feedback Training

Figure 5-3. Initial consonant discrimination. Mean d' (top) and absolute bias (bottom) values for each subject, averaged across the seven initial consonant pairs. Error bars indicate one s.d. from mean. "VF" and "NV" designations on abscissa indicate type of training (vocal-feedback or non-vocal) employed during the corresponding repetition. For each subject, leftmost indicated training strategy (VF or NV) was used during first repetition. Dark and light bars correspond to V-alone and V+T measurements, respectively. In top panel, single (*) and double (**) asterisks indicate p-values ($p < .05$ and $p < .01$, respectively) for one-tailed T-tests comparing V-alone and V+T means for each subject/repetition; dashed line at $d'=1$ corresponds to approximately 70% correct response rate.

pairs under the V-alone and V+T conditions indicate that discrimination enhancement with the addition of tactile cuing was significant at the $p < .01$ level for subjects NH1, NH3, HI1, and HI3 in all repetitions completed, and for subject NH2 in the first repetition. Moreover, V+T discrimination enhancement was significant at the $p < .05$ level for subjects NH2 and HI5 in their second repetitions.

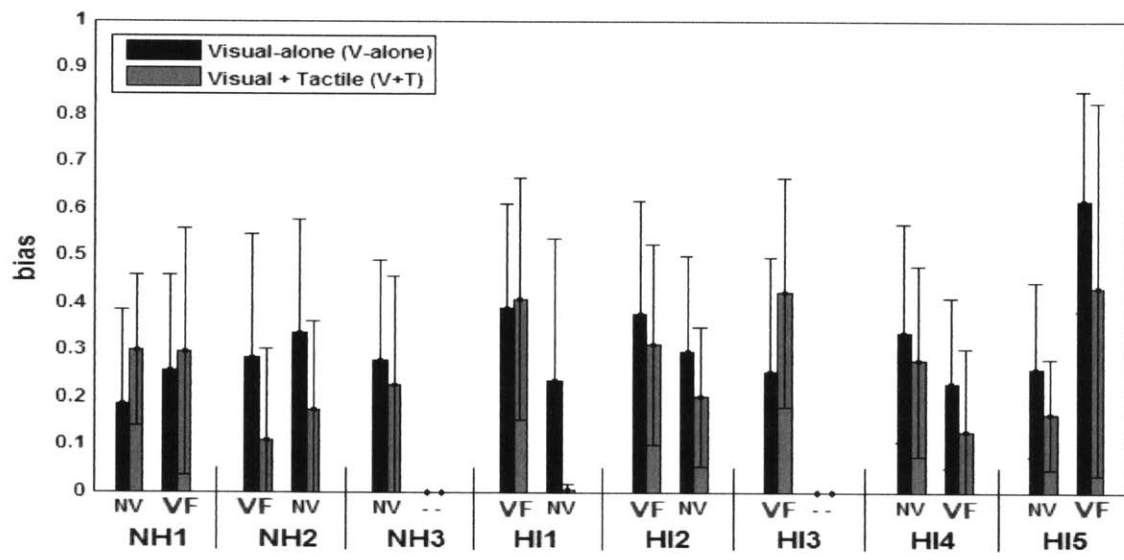
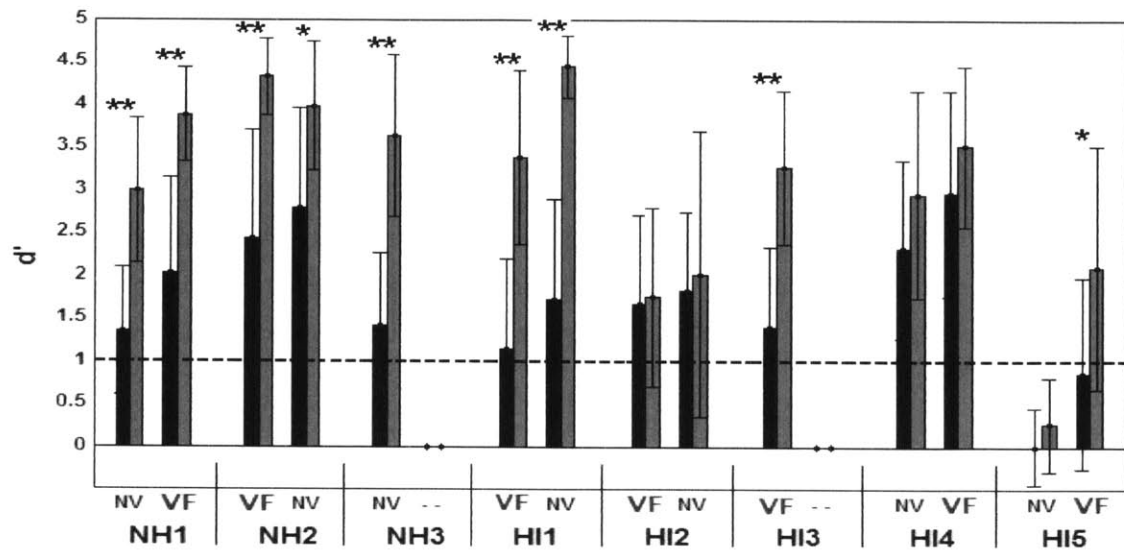
In stark contrast to the other subjects, HI2 and HI4 demonstrated V+T sensitivity levels comparable to those observed under the V-alone condition, in both their first and second repetitions of the experiment. (The fact that these two individuals both exhibited elevated tactile detection thresholds is discussed further below.)

Mean absolute bias in the bottom panels of Figures 5-3 and 5-4 was calculated as the average of the absolute values of the bias for each individual consonant discrimination pair. Mean absolute bias values were generally low, falling below 0.5 in all but one instance. These values suggest that subjects did not exhibit any systematic response bias during pairwise discrimination evaluations for either syllable-initial or syllable-final consonants.

5.3.4 Twelve-Consonant Identification

Figure 5-5 shows percent correct scores for each subject, averaged over 12 or more 50-trial evaluation runs performed in each repetition of the 12-consonant identification experiment. In both the initial consonant (top panel) and final consonant (bottom panel) identification tasks, chance performance was 8.33%. The two tasks were performed simultaneously --- i.e., in each trial, subjects identified both initial and final consonants following a single CVC presentation. Evaluation runs alternated between the V-alone and V+T conditions, results for which are shown by dark and light bars, respectively, in Figure 5-5. One-tailed T-tests were performed to compare V-alone and V+T arcsine-transformed proportion scores for each subject/repetition, and significance of differences is indicated by asterisks in the figure.

All subjects other than NH3 and HI3 performed a second full repetition of the experiment, using the alternative training strategy to that used in the first repetition. Two-tailed T-tests were conducted to compare each subject's V-alone performance in the first and second repetitions, and separately to compare each subject's V+T performance in the two repetitions. In all cases, proportion scores were arcsine-transformed prior to T-test application. The results are summarized in Table 5-5, in which asterisks indicate that a given subject's identification performance for the corresponding consonant position and sensory condition differed significantly between the first and second repetitions. Single (*) and double (**) asterisks denote significance at levels of $p < .05$ and $p < .01$, respectively.



Subject / training-type

NV = Non-vocal Training
 /F = Vocal-feedback Training

Figure 5-4. Final consonant discrimination. Mean d' (top) and absolute bias (bottom) values for each subject, averaged across the seven final consonant pairs. Error bars indicate one s.d. from mean. "VF" and "NV" designations on abscissa indicate type of training employed during the corresponding experimental repetition. For each subject, leftmost indicated training strategy (VF or NV) was used during the first repetition. Dark and light bars correspond to V-alone and V+T measurements, respectively. In top panel, single (*) and double (**) asterisks and dashed line at $d'=1$ are as in Figure 5-3..

Figures 5-7, 5-8, and 5-9 summarize the results of the consonant identification experiment in terms of percent information transfer (%IT) for the individual articulatory features voicing, manner, and place. The %IT provides a measure of the observed covariance of a subject's responses with the stimuli presented, relative to the total information in the stimuli. The %IT for each individual articulatory feature is obtained by considering only the identity of that feature for each stimulus and response in the calculation of transmitted information, thus treating each feature as a separate "communication channel". (For details of %IT calculation, see methods section of this chapter.)

5.3.4.1 CVC-Initial Identification

A three-way analysis of variance (ANOVA) was conducted to test for performance effects of condition (V-alone, V+T), training strategy (NV, VF), and experimental repetition (first, second) in CVC-initial consonant identification. A significant effect was observed for condition, $F(1,21) = 40.64, p < .00001$. No significant effects were demonstrated for training strategy, $F(1,21) = .18, p = .67$, or experimental repetition, $F(1,21) = .68, p = .42$. Moreover, no interaction effects were observed between sensory condition and training strategy, $F(1,21) = .16, p = .69$, between training strategy and repetition, $F(1,21) = .03, p = .87$, or between sensory condition and repetition, $F(1,21) = .13, p = .72$.

As observed in the top panel of Figure 5-5, initial consonant identification performance in the V-alone condition falls roughly in the range of 40-50% accuracy for all subjects. In the V+T condition, performance across subjects varies substantially more, falling roughly between 50-80% accuracy. In all cases, identification performance improved in the V+T condition relative to V-alone. The improvement was found to be significant at a level of $p < .01$ for all but the first repetitions of subjects HI4 and HI5.

The four best performers in the V+T condition (NH1, NH2, NH3, and HI3) each demonstrate approximately 30-40% increased accuracy relative to the V-alone condition, and in each case, the difference in performance under the two conditions is significant at a level of $p < .01$. In their first repetitions of the experiment, subjects NH1 and NH3 trained non-vocally (NV) under the V+T condition, whereas subjects NH2 and HI3 trained with vocal feedback (VF). However, despite having undergone different training strategies, these four individuals performed comparably to one another --- roughly 75-80% accuracy in V+T trials.

Subjects HI1 and HI2, who trained under the vocal strategy in their first repetitions, show improvements of 20% and 10%, respectively, in the V+T condition relative to V-alone (significant at $p < .01$). Subjects HI4 and HI5, who trained under the non-vocal strategy in their first repetitions, show no significant difference in performance under the V-alone and V+T conditions.

The top panel of Figure 5-5 further shows that the six subjects who performed second repetitions all demonstrated significantly better V+T than V-alone performance in initial consonant identification. This includes subjects HI4 and HI5, neither of whom showed a significant effect of tactile cuing in the first repetition, and both of whom trained under the vocal strategy during the second repetition.

Comparing identification performance in the first and second repetitions, Table 5-5 reveals that five out of six subjects (all but HI2) showed significant improvement in the second repetition (at a level $p < .01$) in syllable-initial consonant identification performance under the V+T condition. By contrast, only two subjects, NH1 and HI1, improved significantly in initial consonant identification under the V-alone condition (at levels $p < .05$ and $p < .01$, respectively) between the first and second repetitions.

A two-way, repeated-measures ANOVA was performed to test for effects of training strategy (NV, VF; within-subjects factor) and training order (NV-VF, VF-NV; between-subjects factor) on tactual speech-benefit to CVC-initial consonant identification performance (i.e., the difference between the arcsine-transformed V+T and V-alone scores) among the four hearing-impaired subjects who completed two full procedural repetitions. A significant main effect of training strategy was observed, $F(1, 2) = 26.41$, $p = .036$, $\eta_p^2 = .930$. (Partial eta-squared, η_p^2 , describes the contribution of each factor/interaction to the total variation.) This was qualified by a significant interaction between training strategy and training order, $F(1, 2) = 59.64$, $p = .016$, $\eta_p^2 = .968$. The main effect of training order was non-significant, $F(1, 2) = 1.48$, $p = .348$, $\eta_p^2 = .425$.

Articulatory feature transmission (CVC-initial)

Figure 5-7 relates subjects' reception of the "voicing" feature in CVC-initial and CVC-final positions, characterized as %IT of voicing. In the V-alone condition, %IT of voicing in initial consonant identification (Figure 5-7, top panel, dark bars) is close to zero for all subjects/repetitions. This reflects the paucity of voicing information available solely through visual observation of a speaker's face. In the V+T condition (light bars), %IT ranges from about 4% for HI4's first repetition to about 50% for both NH1 and NH2 in their second repetitions.

In their first repetitions, the four best performers in initial consonant identification (NH1, NH2, NH3 and HI3) exhibit roughly 30-40% improvements in voicing information transmission under V+T relative to V-alone. Of these, NH1 and NH2 performed second repetitions, in which they further increased their relative V+T benefit by an additional 10-15% IT of voicing.

Subjects HI1 and HI2 each showed close to 10% V+T benefit in their first repetitions, in which they had both trained according to the vocal strategy. In their second repetitions, after training non-vocally, HI1 further increased her V+T benefit to about 20% voicing IT, whereas HI2 showed no additional V+T benefit. Subjects HI4 and HI5, who trained non-vocally in the first repetition, showed the poorest reception of voicing information under V+T relative to V-alone (roughly 2% and 5% voicing IT, respectively). However, both HI4 and HI5 improved substantially in voicing reception in the V+T condition after training vocally in their second repetitions, exceeding 15% voicing IT benefit relative to V-alone.

Averaged across all subjects in their final repetitions, initial consonant voicing reception was less than 2% IT in the V-alone condition, and about 31% IT in the V+T condition (Figure 5-10, top-left).

Figure 5-8 relates subjects' reception of the "manner" feature in CVC initial and final positions (top and bottom panels, respectively), characterized as %IT of manner. In the V-alone condition, transmission of manner information for initial consonant identification (Figure 5-8, top panel, dark bars) is roughly in the range of 40-60% for all subjects/repetitions. In the V+T condition (light bars), %IT of manner shows more variability across subjects, ranging from roughly 40-80%. The benefit to manner reception derived from the addition of tactile cuing (V+T condition relative to V-alone) ranges from the nominal benefits observed for the first repetitions of subjects HI2, HI4, and HI5, to greater than 40% improved manner reception in the case of HI3.

In their first repetitions, the four best performers in initial consonant identification (NH1, NH2, NH3 and HI3) exhibit roughly 20-40% (22%, 25%, 32%, and 42%, respectively) improvements in articulatory manner information reception under V+T relative to V-alone. NH1 and NH2, who performed second repetitions, showed small changes (5% increase and 7% decrease, respectively) in V-alone manner reception, and both showed improvement in V+T manner reception (about 11% and 7%, respectively).

Subject HI1 showed similar initial consonant manner reception in the first and second repetitions, exhibiting a consistent V+T benefit of roughly 15% IT. Subjects HI2, HI4, and HI5 each showed nominal V+T benefits to initial consonant manner reception (a few percentage points) in their first repetitions. However, in their second repetitions, all three subjects exhibited roughly 10% improved manner reception in V+T relative to V-alone trials. For HI2, the apparent increase in V+T benefit resulted mainly from a reduction in V-alone manner reception. By contrast, both HI4 and HI5 demonstrated clear improvement in V+T reception of initial consonant manner in the second repetition.

Averaged across all subjects in their final repetitions, initial consonant manner reception was roughly 48% IT in the V-alone condition, and 72% IT in the V+T condition (Figure 5-10, top-center).

Figure 5-9 relates subjects' reception of the "place" feature in CVC initial and final positions (top and bottom panels, respectively), characterized as %IT of place. In the V-alone condition, transmission of articulatory place information for initial consonant identification (Figure 5-9, top panel, dark bars) is roughly in the range of 70-80% for all subjects/repetitions. In the V+T condition (light bars), %IT of place generally falls within a few percentage points of the corresponding V-alone score, suggesting that tactile cuing has little impact on the reception of initial consonant place of articulation. The largest disparity in place reception between the two sensory conditions is the 8% V+T benefit exhibited by subject NH1 in the second repetition.

In their first repetitions, the four best performers in initial consonant identification (NH1, NH2, NH3 and HI3) all exhibit a slight V+T benefit (about 2-3% IT) for articulatory place reception. Of these, NH1 and NH2 performed second repetitions, in which they both demonstrate increased V+T benefit relative to V-alone (8% and 6% IT, respectively). Subjects HI1, HI2, HI4, and HI5 each demonstrate less than 5% IT disparity in place reception between the V-alone and V+T conditions, in both their first and second repetitions. Each exhibits slightly better V-alone than V+T place reception in at least one of their two repetitions (thus breaking with the trend observed for the four best performers).

Averaged across all subjects in their final repetitions, reception of initial consonant articulatory place was about 78% IT in the V-alone condition, and 81% IT in the V+T condition (Figure 5-10, top-right).

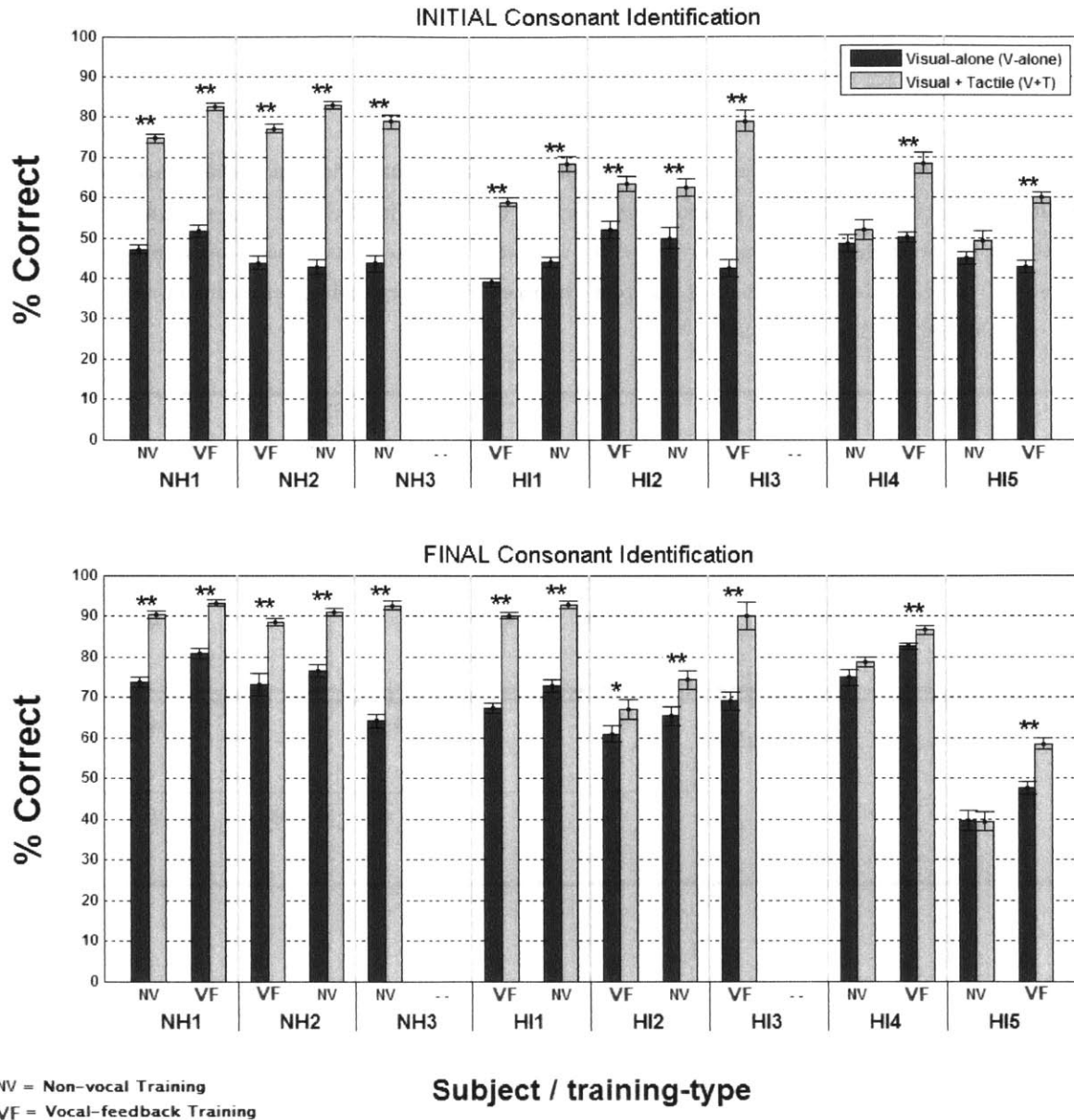


Figure 5-5. Mean percent correct performance in 12-alternative identification of CVC-initial (top panel) and CVC-final (bottom panel) consonants, presented for each subject, in each repetition, under the V-alone and V+T sensory conditions (dark and light bars, respectively). "VF" and "NV" designations on abscissa indicate the type of training (vocal-feedback or non-vocal, respectively) employed during the corresponding experimental repetition. Error bars indicate standard error of the mean (SEM) over multiple 50-trial runs. Single (*) and double (**) asterisks indicate p-values ($p < .05$ and $p < .01$, respectively) for one-tailed T-tests comparing arcsine-transformed V-alone and V+T proportion scores for each subject/repetition.

Table 5-5. Significance of differences between subjects' consonant identification performance under like conditions (i.e., same consonant position and sensory condition) in the first and second procedural repetitions. A two-tailed T-test was performed for each set of arcsine-transformed scores. Single (*) and double () asterisks denote a statistical difference between the corresponding set means at a significance level of $p < .05$ and $p < .01$, respectively.**

		Subjects					
		NH1	NH2	HI1	HI2	HI4	HI5
Syllable-INITIAL Identification	<i>V-alone</i>	*	-	**	-	-	-
	<i>V+T</i>	**	**	**	-	**	**
Syllable-FINAL Identification	<i>V-alone</i>	**	-	**	-	**	**
	<i>V+T</i>	*	-	*	*	**	**

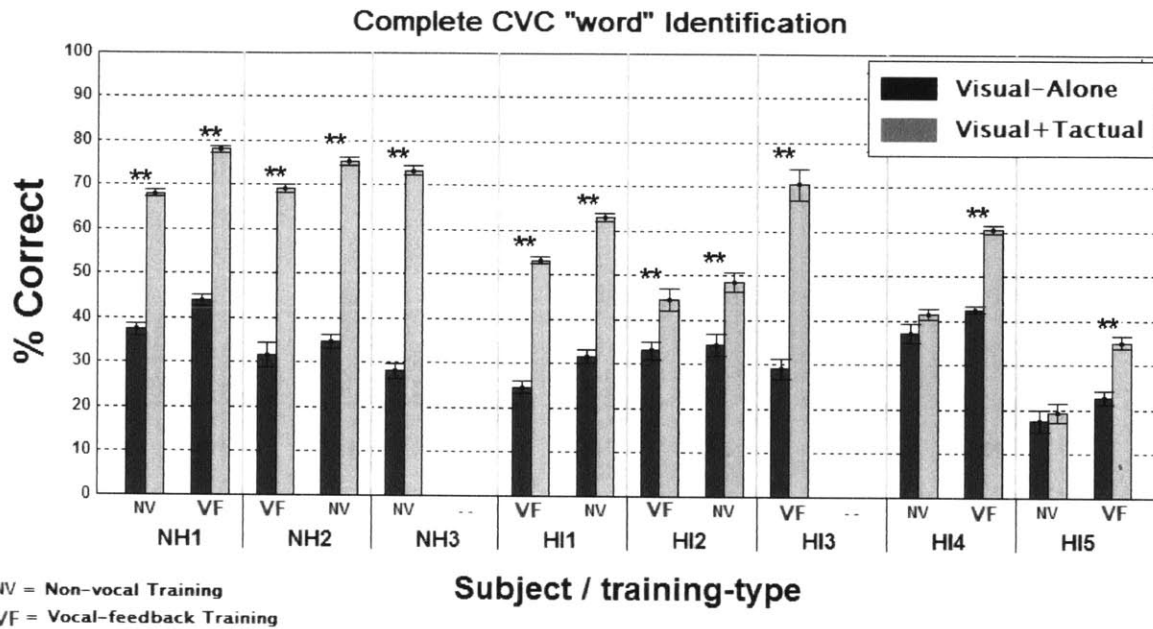


Figure 5-6. Mean percent of correctly identified CVC "words" in 12-alternative identification task. For purposes of this analysis, correct identification of both initial and final CVC consonants constitutes a single correct response (the response is considered incorrect if either consonant is incorrectly identified). Mean data are presented for each subject, in each repetition, under the V-alone and V+T sensory conditions (dark and light bars, respectively). "VF" and "NV" designations on abscissa indicate the type of training (vocal-feedback or non-vocal, respectively) employed during the corresponding experimental repetition. Error bars indicate standard error of the mean (SEM) over multiple 50-trial runs. Double asterisks (**) indicate significance at $p < .01$ for one-tailed T-tests comparing arcsine-transformed V-alone and V+T proportion scores for each subject/repetition.

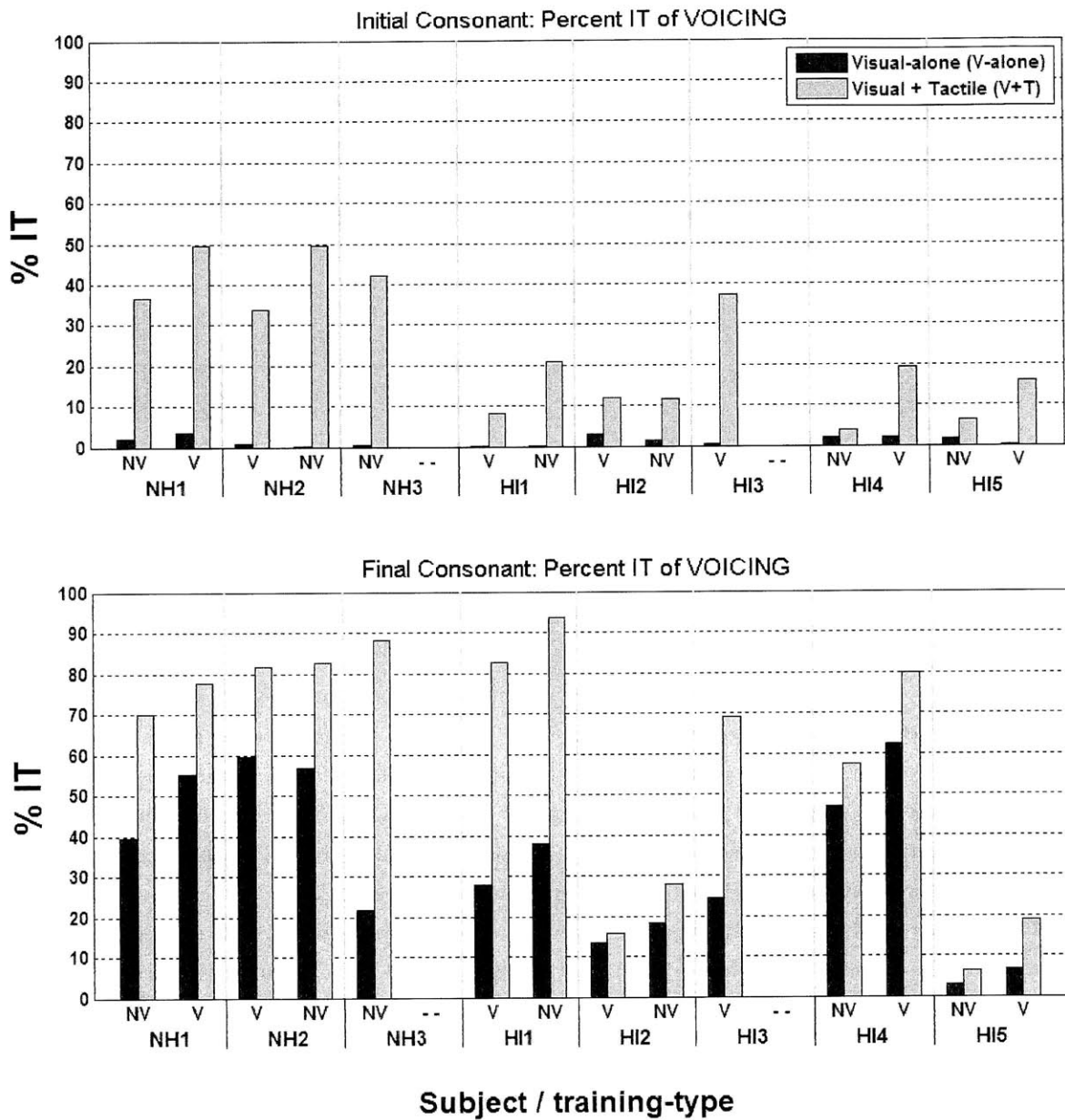


Figure 5-7. Reception of the articulatory feature "voicing" in %IT, presented for each subject, in each repetition, in the 12-alternative identification of CVC-initial (top panel) and CVC-final (bottom panel) consonants, under the V-alone and V+T sensory conditions (dark and light bars, respectively).

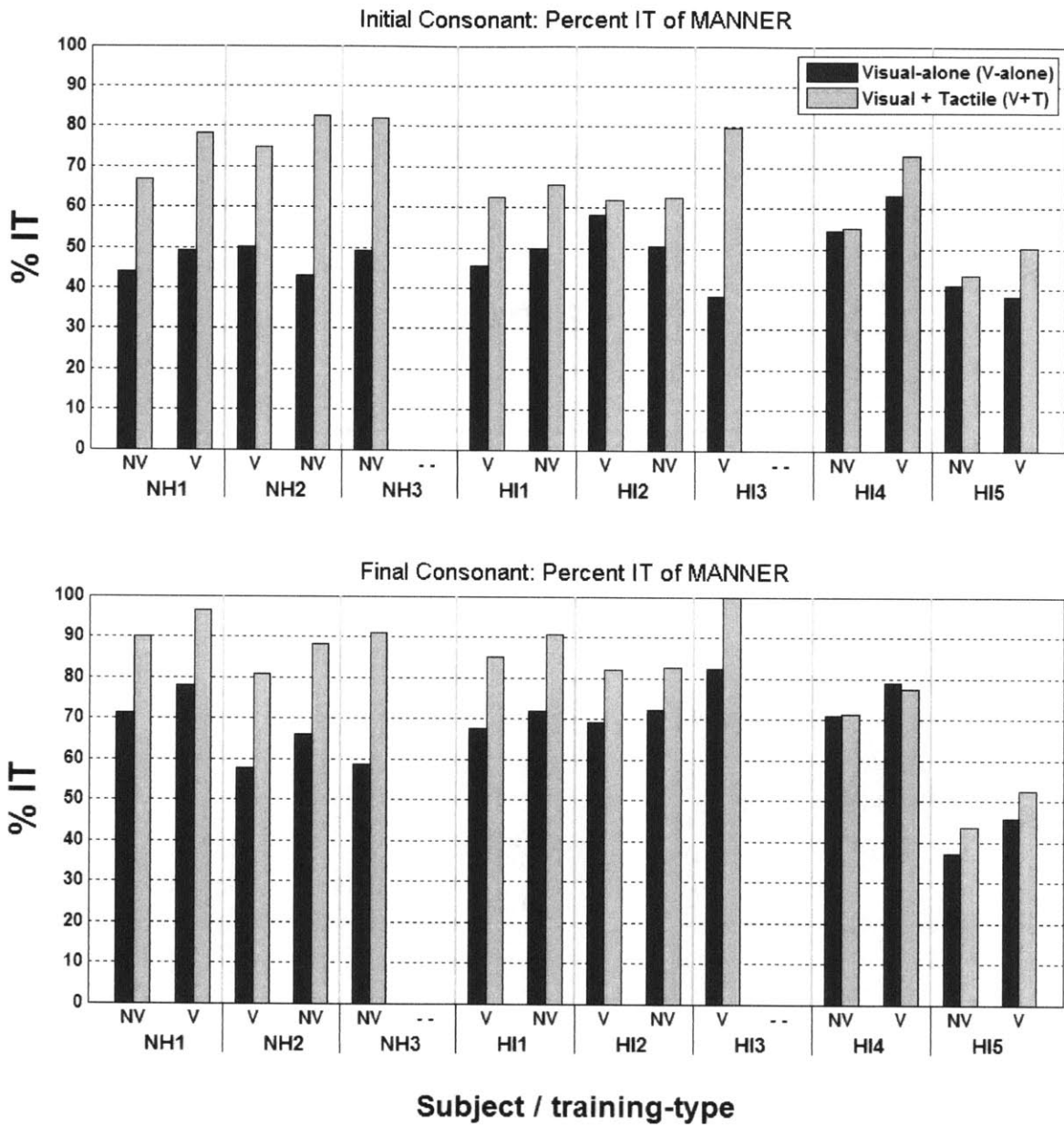


Figure 5-8. Reception of the articulatory feature "manner" in %IT, presented for each subject, in each repetition, in the 12-alternative identification of CVC-initial (top panel) and CVC-final (bottom panel) consonants, under the V-alone and V+T sensory conditions (dark and light bars, respectively).

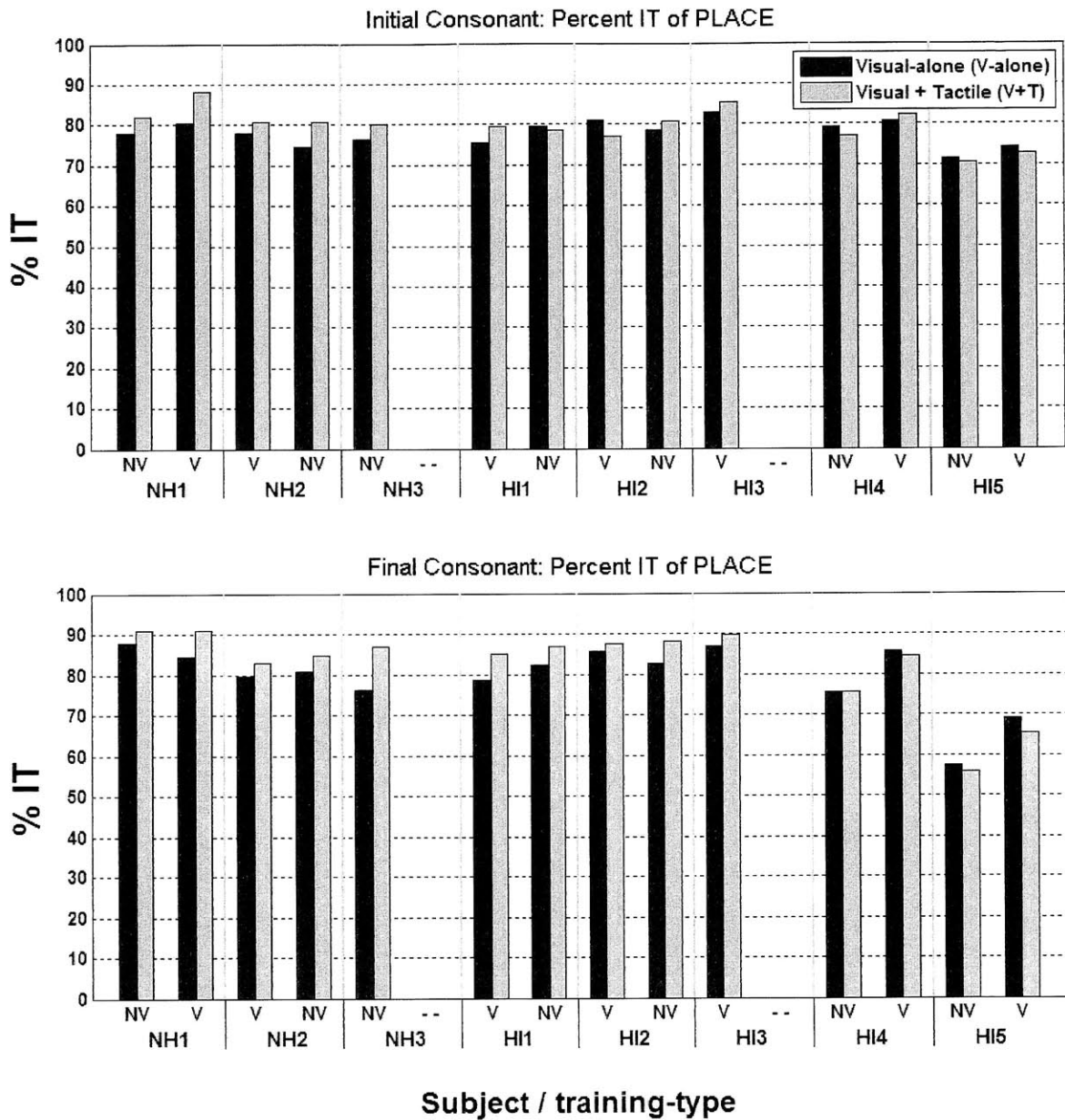


Figure 5-9. Reception of the articulatory feature "place" in %IT, presented for each subject, in each repetition, in the 12-alternative identification of CVC-initial (top panel) and CVC-final (bottom panel) consonants, under the V-alone and V+T sensory conditions (dark and light bars, respectively).

5.3.4.2 CVC-Final Identification

A three-way ANOVA was performed to test for effects of sensory condition (V-alone, V+T), training strategy (NV, VF), and experimental repetition (first, second) on CVC-final consonant identification performance. A significant effect was observed for condition, $F(1,21) = 6.62, p = .02$. No significant effects were demonstrated either for strategy, $F(1,21) = .11, p = .75$, or for repetition, $F(1,21) = .58, p = .46$. Moreover, no interaction effects were noted between condition and strategy, $F(1,21) = .0014, p = .97$, between strategy and repetition, $F(1,21) = .65, p = .43$, or between condition and repetition, $F(1,21) = .04, p = .84$.

The bottom panel of Figure 5-5 reveals that, in the V-alone condition, all subjects other than HI5 demonstrate final consonant identification performance in the range of approximately 60-80% accuracy, which is substantially better than V-alone identification of initial consonants. HI5 performs at about 40% accuracy in the V-alone condition, suggesting a lesser proficiency at discriminating facial cues exploited by the other subjects.

In their first repetitions, subjects' ability to exploit tactual cuing to support syllable-final consonant identification follows a similar pattern to that observed for initial consonants. Subjects NH1, NH2, NH3, HI1, and HI3 all approach or exceed 90% accuracy in the V+T condition. In each case, V+T performance improvement relative to V-alone is found to be significant at a level of $p < .01$, with tactual benefit falling roughly in the range of 20% to 40% (Figure 5-5, bottom). Subject HI2 shows a substantially smaller benefit of about five percentage points in V+T relative to V-alone, which is significant at $p < .05$.

In contrast to the others, subjects HI4 and HI5 do not exhibit significantly different final consonant identification performance in the V-alone and V+T conditions during their first repetitions. Figure 5-5 (bottom) indicates tactual benefit of less than 5% for HI4, and none at all for HI5. Notably, HI4 demonstrates V-alone identification performance comparable to most of the other subjects, whereas HI5 shows only about half the accuracy of HI4 in the V-alone condition.

As was observed for initial consonant identification, the bottom panel of Figure 5-5 indicates that the six subjects who performed second repetitions all showed significantly better V+T than V-alone performance in final consonant identification. The performance of subjects HI4 and HI5 is particularly noteworthy, as neither demonstrated any significant benefit from tactile cuing in the first repetition, during which they had both trained under the non-vocal strategy.

Comparing final consonant identification performance in the first and second repetitions, Table 5-5 indicates that five out of six subjects (all but NH2) showed significant improvement under the V+T condition, but of these, only HI4 and HI5 demonstrated performance improvements at a significance level $p < .01$. Figure 5-5 (bottom) indicates a tactual benefit of about 11% for HI5 in the second repetition, compared to zero benefit in the first repetition. Under the V-alone condition, the performance of all subjects other than NH2 and HI2 improved significantly ($p < .01$) in the second repetition relative to the first.

A two-way, repeated-measures ANOVA was performed to test for effects of training strategy (NV, VF; within-subjects factor) and training order (NV-VF, VF-NV; between-subjects factor) on tactual speech-benefit to CVC-final consonant identification performance (i.e., the difference between the arcsine-transformed V+T and V-alone scores) among the four hearing-impaired subjects who completed two full procedural repetitions. Main effects of training strategy and training order, as well as interaction between them, were all non-significant, $F \leq 1.64$, $p \geq .328$.

Articulatory feature transmission (CVC-final)

Overall, %IT of voicing in final consonant identification (Figure 5-7, bottom panel) is substantially higher than that for initial consonants. In the V-alone condition, %IT of final consonant voicing ranges from about 3% (HI5, first repetition) to greater than 60% (HI4, second repetition). The disparity among subjects reflects the fact that visual observation can potentially provide substantial information about final consonant voicing, but that subjects' ability to exploit this visual information is highly variable. In the V+T condition, %IT of final consonant voicing ranges from about 6% (HI5, first repetition) to nearly 95% (HI1, second repetition). Thus, through combined visual and tactile reception, final consonant voicing may be ascertained nearly perfectly, yet all do not readily acquire the requisite perceptual skills.

In their first repetitions, the five best (overall) performers in final consonant identification (NH1, NH2, NH3, HI1 and HI3) vary substantially in their V-alone reception of final consonant voicing --- NH3, HI1, and HI3 each received in the range of 20-30% of final consonant voicing information via visual input alone, while NH1 and NH2 received approximately 40% and 60%, respectively, via visual input alone. By contrast, in the V+T condition, NH3, HI1, and HI3 show vastly improved voicing reception, ranging about 45-65% higher than V-alone, while NH1 and NH2 improve by roughly 30% and 20%, respectively. For all five subjects, the net result is V+T final consonant voicing information reception roughly in the range of 70-80% (in the first repetition).

Of the five best performers, NH1, NH2, and HI1 performed second repetitions. Relative to their first repetitions, subjects NH1 and HI1 each improved by roughly 10-15% IT in their reception of final consonant voicing under both the V-alone and V+T conditions. In contrast, subject NH2 showed nearly identical final consonant voicing reception in the two repetitions.

Subject HI4 demonstrates impressive V-alone reception of final consonant voicing in the first repetition (~47% IT), but shows only about 10% benefit from the addition of tactile cuing in the V+T condition. In the second repetition, here V-alone and V+T final consonant voicing IT scores increase by about 15% and 22%, respectively. Subjects HI2 and HI5 exhibit relatively poor V-alone reception of final consonant voicing in their first repetitions (13% and 3%, respectively), and both show only nominal benefits of a few percentage points under the V+T condition. In their second repetitions, subjects HI2 and HI5 show slightly improved V-alone final consonant voicing reception, and notably, both benefit more substantially under V+T relative to V-alone (about 9% and 12% enhancement, respectively).

Averaged across all subjects in their final repetitions, final consonant voicing reception was about 36% IT in the V-alone condition, and 68% IT in the V+T condition (Figure 5-10, bottom-left).

Reception of articulatory manner information in the final consonant identification task is depicted in the bottom panel of Figure 5-8. In the V-alone condition, %IT of manner for all subjects falls roughly in the range of 40-80%. In the V+T condition, manner reception ranges from about 44% IT for HI5 in her first repetition to perfect reception (100% IT of manner) for HI3 in her sole repetition.

In their first repetitions, the five best performers in final consonant identification (NH1, NH2, NH3, HI1 and HI3) exhibit roughly 18-32% improvements in manner information reception under the V+T condition relative to V-alone. Of these, NH1, NH2, and HI1 performed second repetitions, in which they show slight increases in manner reception (ranging from 3-7% IT) in the V-alone condition, and V+T benefits comparable to those observed in their first repetitions (roughly 18-22% IT of manner).

Subject HI2 shows roughly 70% manner reception in the V-alone condition in both repetitions. Her reception of final consonant manner in the V+T condition exceeds her V-alone reception by about 12% and 10% IT in her first and second repetitions, respectively. Subject HI4 demonstrates about 70% V-alone manner reception in the first repetition and close to 80% V-alone manner reception in the second repetition. However, she derives no benefit to manner reception from the addition of tactile cuing in

either repetition. Subject HI5 shows the poorest final consonant manner reception by far. Her V-alone manner reception increases from about 38% IT in the first repetition to 46% in the second repetition. In both repetitions, she exhibits approximately 5-6% better manner reception in the V+T condition relative to V-alone.

Averaged across all subjects in their final repetitions, final consonant reception of articulatory manner was about 69% IT in the V-alone condition, and 85% IT in the V+T condition (Figure 5-10, bottom-center).

Reception of articulatory place information in the final consonant identification task is depicted in the bottom panel of Figure 5-9. In both the V-alone and V+T conditions, %IT of place falls roughly in the range of 75-90% for all subjects except HI5, who shows somewhat poorer reception of final consonant place.

Of the five best performers in final consonant identification, subject NH3 demonstrates the most substantial V+T benefit for place reception (roughly 10% IT improvement relative to the V-alone condition). The remaining four "best performers" (NH1, NH2, HI1 and HI3) show V+T benefits of roughly 3-6% IT for place in both first and second repetitions. All five show V+T place reception in the narrow range of about 84-90% IT (in all repetitions).

Subject HI2 demonstrates final consonant place reception comparable to the five best performers in both repetitions. Subject HI4 exhibits no V+T benefit in either repetition, but her place reception under both sensory conditions is similar to the others (roughly 76% and 85% IT in the first and second repetitions, respectively). In contrast to the other seven subjects, HI5 demonstrates only about 57% final consonant place reception in the first repetition, improving by about 10% in the second repetition. Her V+T place reception is about 2% and 4% worse than her V-alone reception in the first and second repetitions, respectively.

Averaged across all subjects in their final repetitions, final consonant place reception was about 81% IT in the V-alone condition, and 85% IT in the V+T condition (Figure 5-10, bottom-right).

Figure 5-10 depicts the relative reception of the three articulatory features voicing, manner, and place, in CVC-initial (top panel) and CVC-final (bottom panel) contexts, averaged across subjects' final repetitions. For all three features, in both phonemic contexts, reception in the V+T sensory condition

(light bars) exceeds that in the V-alone condition (dark bars). The relative benefit of tactile cuing to reception of consonant voicing is approximately 30% IT for both initial and final consonants. Tactile supplementation of lipreading also imparts substantial benefit to the reception of articulatory manner, improving manner transmission by roughly 24% and 16% for initial and final consonants, respectively. In contrast to receptive enhancement associated with voicing and manner, the current tactile cuing scheme appears to convey relatively little information about place of consonant articulation, contributing only about 3-4% to place reception through lipreading alone in both syllable-initial and final contexts.

Complete CVC "Word" Identification

Correct "*word*" identification, as defined for this analysis, requires correct identification of both initial and final CVC consonants. Since the vowels /a/, /i/ and /u/ are visually distinct from one another, correct identification of the vowel is assumed. The response is considered incorrect if either consonant is incorrectly identified.

Figure 5-6 presents the mean proportion of correctly identified CVC *words* in 12-alternative identification task. Mean data are presented for each subject, in each repetition, under the V-alone and V+T sensory conditions (dark and light bars, respectively). "VF" and "NV" designations on abscissa indicate the type of training (vocal-feedback or non-vocal, respectively) employed during the corresponding experimental repetition. Error bars indicate standard error of the mean (SEM) over multiple 50-trial runs. Double asterisks (**) indicate significance at $p < .01$ for one-tailed T-tests comparing arcsine-transformed V-alone and V+T proportion scores for each subject/repetition.

As one might expect, mean percent correct scores for complete *word* identification are all lower than the corresponding scores for initial consonant and final consonant identification individually. V-alone *word* scores range from about 18% to 43%. V+T *word* scores range from about 20% to 78%. The difference between V-alone and V+T performance is significant at $p < .01$ for all subjects in all repetitions, with the exceptions of the first repetition scores for subjects HI4 and HI5.

The four best performers in the V+T condition (NH1, NH2, NH3, and HI3) demonstrate tactual speech-benefit ranging from about 30-45% increased *word* identification accuracy. In their first repetitions of the experiment, subjects NH1 and NH3 trained non-vocally (NV), whereas subjects NH2 and HI3 trained with vocal feedback (VF) under the V+T condition. We find that these four individuals performed comparably to one another, identifying CVC words with mean V+T accuracies in the range of 68-73%,

despite differences in training strategy. NH3 and HI3 scored lowest among the four subjects under the V-alone condition, but compensate under the V+T condition by achieving greater tactual speech-benefit (45% and 41%, respectively). Thus, the choice of training strategy appears inconsequential for these top performers, who learned quickly to exploit tactual speech cues.

Subjects NH1 and NH2 both performed second procedural repetitions, training according to the VF and NV protocols, respectively. Subject NH1 scored 44% correct under V-alone and 78% correct under V+T, which were improvements of six and ten percentage points, respectively, relative to the first repetition. Subject NH2 scored 35% correct under V-alone and 75% correct under V+T, which were improvements of three and six percentage points, respectively, relative to the first repetition.

The four remaining subjects are all profoundly deaf individuals, who completed two full procedural repetitions. HI1 and HI2 trained according to the VF strategy in their first repetitions and the NV strategy in their second repetitions. In their first repetition, complete *word* scores for HI1 and HI2, respectively, were 25% and 33% under the V-alone condition and 53% and 45% under the V+T condition. In their second repetition, complete *word* scores for HI1 and HI2, respectively, were 32% and 35% under the V-alone condition and 63% and 48% under the V+T condition. For both subjects, the difference between V-alone and V+T performance is significant ($p < .01$) for both repetitions.

HI4 and HI5 trained according to the NV strategy in their first repetitions and the VF strategy in their second repetitions. In their first repetition, complete *word* scores for HI4 and HI5, respectively, were 37% and 17% under the V-alone condition and 42% and 20% under the V+T condition. In their second repetition, complete *word* scores for HI4 and HI5, respectively, were 43% and 23% under the V-alone condition and 61% and 35% under the V+T condition. For both subjects, the difference between V-alone and V+T performance is significant ($p < .01$) only for their second repetitions.

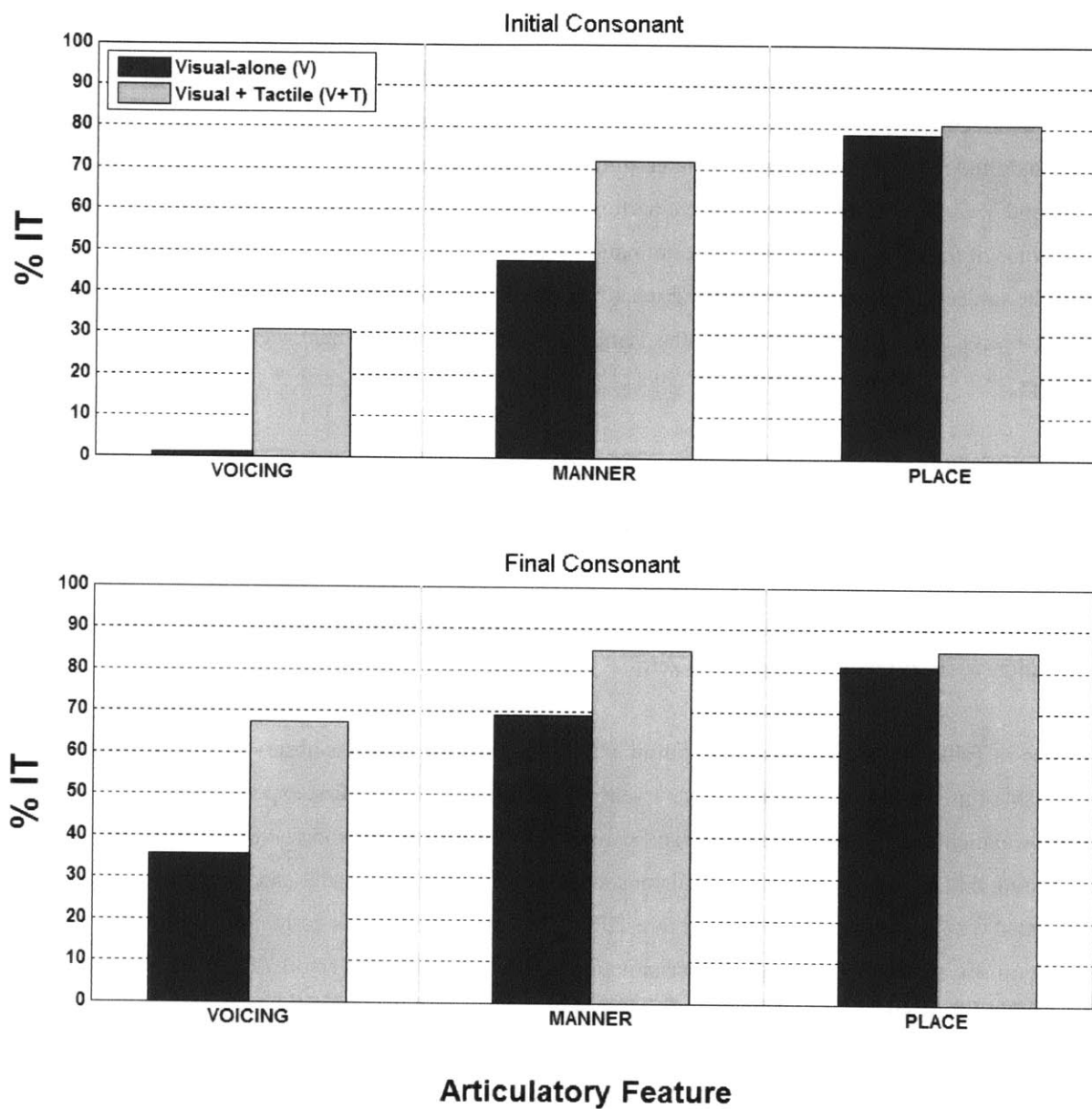


Figure 5-10. Mean performance across subjects in %IT for the three articulatory features voicing (left), manner (center), and place (right), in 12-alternative identification of CVC-initial (top panel) and CVC-final (bottom panel) consonants, under the V-alone and V+T sensory conditions (dark and light bars, respectively). Mean score calculations included evaluation data from each subject's final repetition of the identification experiment.

5.3.5 Correlation of Psychophysical Measurements with Tactual Speech Reception

Correlation analyses were performed in order to examine possible correlations between tactual temporal order resolution (from section 5.3.2) and pairwise discrimination of consonants, in the presence and absence of tactile cuing. Table 5-6 presents correlation coefficients between tactual temporal onset-order thresholds and mean d' values across syllable-initial and syllable-final consonant pairs, for conditions V-alone and V+T. Correlation analyses were performed twice, first using d' values from the subjects' first repetitions of the pairwise consonant discrimination evaluation, and then a second time using d' values from the subjects' final repetitions (second repetition for all but NH3). Measurements from seven out of eight subjects, those who participated in the Temporal Onset-Order experiment, were included these analyses.

In these as well as subsequent correlation analyses, the significance of the sample-derived correlation coefficient, r , was tested using the formula $t = r \sqrt{\frac{n-2}{1-r^2}}$, which approximates Student's t distribution with $(n - 2)$ degrees of freedom, where n is the sample size. The corresponding p -value reflects the probability of observing a correlation coefficient r of equal or greater magnitude under the null hypothesis that the population-wide correlation ρ is equal to zero.

As seen in Table 5-6, the strongest correlation with onset-order thresholds is observed for discrimination of syllable-initial consonants under the V+T sensory condition during the first experimental repetition. The correlation coefficient, $r = -0.976$ (significant at $p < .01$) indicates a strong inverse correlation, suggesting that higher sensitivity to initial consonant contrasts during tactually-enhanced lipreading is associated with lower tactual onset-order thresholds. The magnitude of this correlation between temporal resolution and V+T initial consonant discrimination is substantially lower, $r = -0.774$, when considering the subjects' final repetitions (significant at $p < .05$).

More generally, subjects' continued experience and practice between the first and final repetitions appears to have had the effect of shifting the correlation between tactual temporal resolution and consonant discrimination performance in the positive direction, for both the V-alone and V+T conditions and for discrimination of both initial and final consonant contrasts. In the cases of V+T syllable-initial discrimination and both V-alone and V+T syllable-final discrimination, the positive shift in correlation coefficients between the first and final repetitions is effectively a reduction in the overall magnitude of the correlation. However, in the case of V-alone syllable-initial discrimination, the correlation coefficient changes from -0.195 for the subjects' first repetitions to 0.532 for their final repetitions. This is the only

positive correlation coefficient in Table 5-6 and suggests the emergence (with practice) of a modest positive correlation between high onset-order thresholds and V-alone initial consonant discrimination performance.

Further correlation analyses were performed to examine correlations between observed "tactile speech-benefit" in the consonant identification task and both tactual sensitivity and temporal order resolution. For purposes of these analyses, the "tactile speech-benefit" for each subject was characterized as the difference between the V+T and V-alone mean arcsine-transformed scores. The resulting correlation coefficients are shown in Table 5-7.

The first nine data columns of Table 5-7 present correlation coefficients between subjects' observed tactual speech-benefit and their tactual detection thresholds measured at the index finger, middle finger, and thumb of the left hand, at vibratory frequencies of 10 Hz, 50 Hz, and 200 Hz (from section 5.3.1). All correlation coefficients are negative, which suggests that overall, higher detection thresholds (lower vibrotactile sensitivity) are correlated with reduced benefit of tactile cuing to speech reception. The strongest and most consistent correlations across repetitions and consonant context are observed for the index finger and thumb at 200 Hz and for the middle finger at 10 Hz and 50 Hz.

Identification of consonants in syllable-initial context correlates most strongly with 10 Hz sensitivity at the middle finger and 200 Hz sensitivity at the thumb. In the former case, correlation coefficients of -0.905 and -0.891 are observed for the first and final repetitions, respectively (significant at $p < .01$). In the latter case, correlation coefficients of -0.937 and -0.965 are observed for the first and final repetitions, respectively (significant at $p < .01$).

By contrast, identification of consonants in syllable-final context correlates most strongly with 200 Hz sensitivity at the index finger, for which the correlation coefficient is -0.867 for the subjects' first repetitions (significant at $p < .05$) and -0.940 for the subjects' final repetitions (significant at $p < .01$).

The final column of Table 5-7 presents correlation coefficients between subjects' tactual speech-benefit and their tactual temporal onset-order thresholds. The pattern of correlation is consistent with that observed for pairwise discrimination (Table 5-6) in that the strongest (negative) correlation is observed in connection with tactile facilitation of syllable-initial consonant reception, particularly for subjects' first experimental repetitions, where the correlation coefficient between onset-order thresholds and tactual

benefit is -0.895 (significant at $p < .01$). For initial consonant identification in the subjects' final repetitions, the correlation coefficient (-0.742) falls just short of significance (at the $p < .05$ level).

Table 5-6. Correlation coefficients between tactual temporal onset-order thresholds and mean d' across syllable-initial and syllable-final consonant pairs, for conditions V-alone and V+T, in subjects' first and final repetitions of the pairwise consonant discrimination evaluation (for subject NH3, first repetition was final repetition). Light and dark shading indicate that correlation coefficient differs significantly from zero at levels $p < .05$ and $p < .01$, respectively. Note: Subject HI3 did not participate in Temporal Onset-Order experiment and was excluded from correlation analysis.

		FIRST Repetition	FINAL Repetition
Syllable-INITIAL Discrimination	<i>V-alone</i>	- 0.195	0.532
	<i>V+T</i>	- 0.976	- 0.774
Syllable-FINAL Discrimination	<i>V-alone</i>	- 0.462	- 0.169
	<i>V+T</i>	- 0.678	- 0.320

Table 5-7. Correlation coefficients between tactual thresholds and differences between V+T and V-alone mean arcsine-transformed scores in 12-alternative identification of CVC-initial and CVC-final consonants, in subjects' first and final repetitions. For the first nine data columns, correlation analysis was performed with subjects' tactual detection thresholds at the fingers and vibratory frequencies indicated in the top two rows. For the final column, correlation analysis was performed with subjects' tactual temporal onset-order thresholds. Light and dark shading indicate that correlation coefficient differs significantly from zero at levels $p < .05$ and $p < .01$, respectively.

		Index			Middle			Thumb			Onset-Order
		10 Hz	50 Hz	200 Hz	10 Hz	50 Hz	200 Hz	10 Hz	50 Hz	200 Hz	
Syllable-INITIAL Identification	<i>FIRST Repetition</i>	-0.780	-0.608	-0.810	-0.905	-0.937	-0.676	-0.749	-0.506	-0.937	-0.895
	<i>FINAL Repetition</i>	-0.644	-0.542	-0.742	-0.891	-0.806	-0.820	-0.661	-0.314	-0.965	-0.742
Syllable-FINAL Identification	<i>FIRST Repetition</i>	-0.541	-0.301	-0.867	-0.775	-0.836	-0.725	-0.582	-0.439	-0.815	-0.670
	<i>FINAL Repetition</i>	-0.584	-0.343	-0.940	-0.808	-0.788	-0.763	-0.472	-0.247	-0.790	-0.619

5.4 Discussion

5.4.1 Tactual Sensitivity

Tactual detection thresholds measured at the left thumb, index, and middle fingers, using sinusoidal vibrotactile stimuli at frequencies of 10 Hz, 50 Hz, and 200 Hz. Overall, detection thresholds (plotted in Figure 5-1) follow the expected pattern of increasing sensitivity as frequency increases from 10 Hz to 200 Hz. Mean values (summarized for each finger-frequency combination in Table 5-3) are consistent with measurements reported in Chapter 3 of this dissertation (and published in Moallem et al., 2010), as well as vibrotactile thresholds reported previously in the literature (e.g., Lamore et al., 1986; Rabinowitz et al., 1987; Tan, 1996; Yuan et al., 2005a).

Despite overall agreement of Figure 5-1 with normative data, subjects HI2 and HI4 exhibit distinct patterns of threshold elevation, suggesting individual reductions in tactual sensitivity that warrant further consideration. Subject HI2 (unfilled square marker in Fig. 5-1) demonstrates sensitivity thresholds within about one standard deviation (s.d.) of the group mean for all three fingers at frequencies of 10 Hz and 50 Hz. However, at 200 Hz, her thresholds for all three fingers are more than one s.d. above their respective group means. In particular, the middle finger 200 Hz threshold for HI2 is just under 0 dB re 1 μ m peak displacement. By comparison, Table 5-3 indicates that the group mean (\pm s.d.) threshold for the middle finger at 200 Hz is -14.7 (\pm 7.7) dB re 1 μ m peak. Moreover, the next highest threshold after that of HI2 is about -12 dB re 1 μ m peak displacement.

Vibrotactile threshold elevation in the frequency range of 200 Hz, particularly at the left middle finger (as demonstrated by HI2), could substantially interfere with one's ability to exploit tactual speech cues characteristic of the vibrotactile encoding strategy adopted for this study. As described above in Chapter 4, the middle finger received 50-400 Hz band-pass filtered input from the low-frequency speech band. The primary role of this channel is to convey information about voicing, selectively presenting content in range of human vocal fundamental frequencies, which fall roughly between 80 Hz and 400 Hz. In the current study, CVC speech stimuli consisted of the recorded utterances of two female speakers, both of whom had an average voice fundamental frequency close to 200 Hz. Thus, vibrations close to 200 Hz presented to the middle finger provide crucial voicing information, required by subjects for successful utilization of tactile speech cues.

Subject HI4 (unfilled triangle marker in Fig. 5-1) demonstrates detection threshold elevation at the left index finger, at all three frequencies tested. At 10 Hz, HI4 has a threshold of 36 dB re 1 μ m peak displacement at the index finger, whereas the group mean (\pm s.d.) threshold is 26.8 ± 5.2 dB re 1 μ m peak. At 50 Hz, her index finger threshold is about 19.3 dB re 1 μ m peak, compared to the group mean (\pm s.d.) threshold of 11.4 ± 4.3 dB. At 200 Hz, HI4 has a index finger threshold of about -6 dB re 1 μ m peak, comparable to that of subject HI2 (discussed above). By comparison, the group mean (\pm s.d.) index finger threshold at 200 Hz is -16.8 ± 8.7 dB re 1 μ m peak displacement.

Substantially reduced vibrotactile sensitivity at the index finger across all frequencies tested may be expected to impair substantially the ability of HI4 to acquire tactual speech reception in this study, although in a manner somewhat distinct from the frequency-specific deficit exhibited by HI2. Under the present tactile speech-encoding scheme, the index finger receives input derived from the high-frequency speech band, although the frequency content of the vibrotactile signal is concentrated in the range of approximately 5 Hz to 400 Hz. In contrast to the middle finger, the index finger receives substantial stimulation at frequencies well below the range of voicing fundamental frequencies. These low-frequency tactual signals convey substantial information about high-frequency acoustic bursts and sustained friction, which supports discrimination of articulatory manner and plays a key role in many consonant-voicing distinctions.

Indeed, as observed in the previous section, subjects HI2 and HI4 demonstrate very limited benefits from tactile speech cuing, relative to most other participants in this study as well as that of Yuan et al. (2005b). The implications of their performance in evaluating the current tactual speech transduction and training strategies are addressed below.

5.4.2 Tactual Onset-Order Discrimination

Discrimination of temporal order among tactual stimuli plays a key role in the effective reception of tactually encoded acoustic speech cues in this study. Onset-order resolution holds particular interest, as it constrains tactual reception of acoustic envelope onset asynchrony (EOA) cues, which can substantially support initial consonant voicing discrimination in the absence of audition. Whereas an effective lipreader can at times discern final consonant voicing from articulatory cues visible on the speaker's face, such visible cues are almost entirely absent during initial consonant voicing. The effects of articulatory cue visibility on the discrimination of consonant voicing pairs are well illustrated in Figure 5-2 (either panel), where the seven leftmost and seven rightmost bar pairs reflect initial and final consonant voicing, respectively. Note that all consonant pairs labeled on the abscissa, other than "TSi" and "TSf" are

consonant voicing pairs, their articulation being distinguished only by the voicing feature. The dark bars indicate average subject sensitivity to each voicing contrast in the V-alone condition. They reveal that, during unsupplemented lipreading, d' falls close to zero for the six consonant voicing pairs in initial position, indicating that these initial consonant voicing is discriminated at near-chance levels. By contrast, d' values in the V-alone condition are substantially higher for the six consonant voicing pairs in final position, reflecting the fact that subjects are able to exploit visual cues from the speakers' faces to discriminate final consonant voicing with moderate success.

In assessing subjects' basic psychophysical performance, correlations between onset-order resolution and initial consonant voicing discrimination were thus deemed to hold particular relevance among other possible temporal processing measures. Note, however, that the results of Chapter 3 demonstrate a robust correlation across subjects between onset- and offset-order thresholds. Although offset-order thresholds are substantially higher than onset-order thresholds overall, the two scores tend to scale proportionally.

Tactual onset-order thresholds were measured with an adaptive 1I-2AFC procedure, varying the relative onset timing of stimuli at the left index and middle fingers. Both stimuli were 200 Hz sinusoidal vibratory pulses with relatively long durations and simultaneous offsets (the longer stimulus duration was 800 ms). This protocol was chosen for simplicity and brevity, which made it possible to complete temporal threshold measurements in a single experimental session. By contrast, the experimental protocol employed for temporal threshold measurement in Chapter 3 of this dissertation involved a non-adaptive procedure, in which 50 Hz and 250 Hz stimuli were delivered to the left thumb and index finger, respectively. The protocol of Chapter 3 was chosen in accordance with the methods employed by Yuan et al. (2005), which permitted far more detailed examination of the effects of relative stimulus levels and durations on temporal order judgments. Measurements for each subject were conducted in multiple experimental sessions over the course of several weeks.

The onset-order experiment described in this section yielded a mean threshold of 175.7 ms across seven subjects, with individual thresholds ranging from 52.8 ms to 294.1 ms (see Table 5-4). The adaptive paradigm employed in the current study is expected to converge on a stimulus onset-asynchrony (SOA) value at which the subject exhibits roughly 71% correct performance (Levitt, 1971). The non-adaptive protocol of Chapter 3 yielded a mean onset-order threshold of 73.9 ms across 14 subjects, with individual thresholds ranging from 41 ms to 165 ms. In that study, threshold was defined as the interpolated $|SOA|$ value at unity d' , corresponding to approximately 70% correct performance. Clearly, the results of the

two experiments differ substantially, despite having been carried out by the same experimenter, using the same tactual display equipment.

Notably, one experimental subject participated in both the current study and that of Chapter 3, and thus provides a basis for direct comparison of threshold measurement under the two protocols. Subject NH3 (the experimenter and author of the current study), who was found to have an onset-order threshold of 52.8 ms, is the same individual referred to in Chapter 3 as “subject NH2”, where his onset-order threshold measured 41 ms. In both studies, this subject exhibited the lowest overall onset-order threshold. The current measurement exceeds the previous measurement by a factor of approximately 1.3, which might be considered as a putative adjustment factor in comparing onset-order measurements in the two studies. Adjustment of the Chapter 3 mean (73.9 ms) and highest (165 ms) thresholds by a factor of 1.3 yields 96.1 ms and 214.5 ms, respectively. These values still fall well below the current study’s mean and highest thresholds of 175.7 ms and 294.1 ms, respectively. The 1.3-fold increase in threshold measured for subject NH3 in the current relative to the earlier study suggests that thresholds measured under the two protocols are not directly interchangeable. On the other hand, the fact that this same subject exhibited the lowest threshold in both studies is consistent with the a priori expectation that the two measurements should be highly correlated. Considering that the onset-order threshold measured for subject NH3 in the current study exceeded his Chapter 3 by only 30%, the increase of more than 75% in the observed group mean most likely reflects a genuine difference in tactual temporal order sensitivity between the two groups of participants.

With that being said, the disparity between the two sets of measurements may, to some extent, be attributable to differences in one or more experimental parameters, including the use of SOA adaptation, the choice of vibratory frequency at each site, the choice of thumb versus middle finger as stimulation site, and the presence or absence of stimulus level and duration roving.

In contrast to the constant stimuli procedure of Chapter 3, the adaptive procedure used in the current study focuses each run of testing on stimuli close to the subject’s perceptual threshold, converging upon a pre-determined correct response rate (Levitt, 1971). One potential caveat is that this approach assumes nominal response bias, which cannot be confirmed directly. However, signal detection analyses in the earlier temporal order experiment and the pairwise consonant discrimination of the current study, both of which involved similar 1I-2AFC discriminations, have consistently indicated minimal response bias across subjects. Moreover, Taylor et al. (1983) compared detection performance using adaptive and

constant stimuli procedures, and they report that subjects exhibit less sensitivity fluctuation in the adaptive procedure, owing in part to reduced sequential dependency.

Another possible source of variation is the use of 200 Hz vibrotactile stimuli at each site in the current study, in contrast to the use of 50 Hz at the thumb and 250 Hz at the index finger in the earlier study. In a set of “corollary measurements”, Yuan et al. (2005a) found no significant difference in onset-order thresholds measured using 50 Hz at the thumb and 250 Hz at the index finger when compared with measurements using the same frequency stimulation at both fingers. However, the same group (Yuan et al., 2006) later performed a more extensive study of stimulation site and frequency effects upon onset-order judgments, in which they report a significant 5-10 ms elevation of thresholds when the thumb and index finger are stimulated with the same frequency, as compared with when the two stimulation frequencies are separated by several octaves. Thus, the use of 200 Hz stimuli at both sites in the current study may, to some extent, account for the observed disparity with the earlier set of measurements.

The pairing of index finger with the thumb in the earlier study and with the middle finger in the current study cannot readily be ruled out as a source of variation. The literature concerning tactile resolution of temporal order among sustained vibratory stimuli does not address this issue directly. Heming and Brown (2005) demonstrated that thresholds for the perception of simultaneity among punctate tactile stimuli are significantly higher when the stimuli are delivered to index fingers on opposite hands, as compared to when they are delivered to the index and middle fingers of one hand. They argue, however, that the increased thresholds observed in the bimanual task reflect delays associated with interhemispheric processing. One would not expect such delays to factor into temporal processing of tactile stimuli delivered to ipsilateral digits.

As discussed in Chapter 3, stimulus duration did not have any significant effect on the tactual discrimination of onset-order, despite the fact that paired stimuli were independently varied among durations ranging from 50 ms to 800 ms. Rather, duration was found to influence offset-order discrimination selectively. It is thus unlikely that the consistent pairing of relatively long duration stimuli in the current study (800 ms total duration) could have contributed to onset-order threshold elevation.

By contrast, the results of Chapter 3 indicate that relative stimulus level can substantially influence tactual onset-order judgments. In light of this finding, care was taken in the current study to match the amplitudes of 200 Hz vibrotactile stimuli delivered by the two tactual stimulators used in adaptive SOA discrimination procedure. The two stimuli were consistently presented 15 dB re 1 μ m peak displacement.

As observed in the rightmost plot of Figure 5-1, 200 Hz thresholds at the left index and middle fingers of most subjects differed only slightly. Thus, the stimuli used to examine onset-order discrimination in the current study were presented to the two fingers at similar sensation levels, in which case the results of Chapter 3 suggest that sensitivity to SOA should consistently be at its highest. Stimulus level settings cannot explain the disparity between SOA thresholds measured in the two studies.

5.4.3 A Visual Cue for Final Consonant Voicing

Through lipreading alone, corresponding voiced and unvoiced consonants (e.g., /b/-/p/) in CVC-initial context proved virtually indistinguishable to subjects. This is evidenced by the near-zero cross-subject average d' values observed under the V-alone condition for the six voicing-contrast consonant pairs on the left side of Figure 5-2, as well as the near-zero average d' values observed for all subjects under the V-alone condition in the top panel of Figure 5-3.

However, most subjects performed well above chance when discriminating consonant voicing in CVC-final context. Although discrimination of the six CVC-final voicing-contrast consonant pairs varied substantially among subjects, the cross-subject average d' scores under the V-alone condition are all greater than or equal to unity (corresponding to roughly 70% accuracy), and for the subjects' final repetitions (Figure 5-2, bottom-right), the average V-alone d' scores for the consonant pairs /t/-/d/, /p/-/b/, and /f/-/v/ all exceeded $d' = 2$ (corresponding to roughly 85% accuracy).

One possible explanation for the discrepancy between CVC-initial and -final consonant voicing discriminability is that the duration of the vowel (or of the CVC utterance as a whole) is providing a visible cue that is correlated with CVC-final consonant voicing. In auditory speech perception, vowel duration has long been known to provide an acoustic cue for final consonant voicing (e.g., House and Fairbanks, 1953; Denes, 1955; House, 1961; Raphael, 1972; Mack, 1982; Luce and Charles-Luce, 1985). Two primary lines of evidence for such an acoustic are commonly cited. First, acoustic analysis of utterances differing only in syllable-final consonant voicing has consistently revealed longer vowel durations in conjunction with voiced final consonants. Second, in pairwise consonant discrimination testing with synthesized utterances and modified speech recordings, it has been found that increasing vowel duration increases the likelihood of a listener identifying the final consonant as voiced, whereas decreasing vowel duration increases the likelihood of a listener identifying the final consonant as unvoiced.

Although consonant voicing per se is not visible on the face of a speaker, the duration of a vowel in CVC context is visually cued by the consonants that bound it, assuming that the consonants have visible attributes. For example, in distinguishing visually between the nonsense syllables /b a p/ and /b a b/ produced by the same speaker, if the two utterances differ in vowel length, a perceptive observer may note a corresponding difference in the temporal separation of the initial and final consonant plosive bursts. To the extent that the vowel duration does in fact correlate with final consonant voicing (as the literature suggests), subjects in this study could have learned to exploit visual indicators of vowel duration to discriminate between utterances with voiced and unvoiced final consonants. Recognition of this visual cue would be made simpler by the fact that all speech stimuli were produced in an intentionally unexpressive manner, by one of two speakers, and subjects received correct-answer feedback throughout pairwise discrimination training.

5.4.4 Tactually-Assisted Consonant Recognition

The overall pattern of consonant discrimination observed in Figure 5-2 indicates that subjects benefited from tactual cues in discriminating all seven consonant pairs. Tactual benefit to consonant discrimination was substantially greater in syllable-initial position than in syllable-final position, as is clear from the results of T-tests comparing the V-alone and V+T mean d' scores for each consonant pair, particularly in the top panel of Figure 5-2.

Improvement in performance between the top and bottom panels of Figure 5-2 can be ascribed largely to inclusion in the bottom panel averages of measurements from second procedural repetitions completed by six of the eight subjects. By their second repetitions, those six individuals had substantially more practice in the various consonant recognition tasks. Note that, for the consonant pair /f-v/ in both syllable-initial and final positions, as well as the consonant pair /ʃ-ʒ/ in syllable-final position, significant differences are observed between V-alone and V+T performance in the bottom panel only. However, as previously mentioned, two-tailed T-tests reveal no significant differences between corresponding scores in the top and bottom panels of Figure 5-2.

The observation that tactual cuing more consistently benefits consonant discrimination in initial position can be largely understood by considering subjects' performance in the V-alone condition. Specifically, V-alone performance indicates that all subjects perform at or near chance levels when attempting to discriminate the seven feature-contrast consonant pairs in initial position using visual-facial cues only. However, such is clearly not the case for V-alone discrimination of final consonant voicing pairs. Rather,

subjects vary substantially in their ability to derive final consonant voicing cues via visual speechreading. (The visual cue supporting final consonant voicing discrimination is discussed above in section 5.4.3.)

Unlike the voicing contrast pairs, the manner contrast pair /s/-/t/ is not discriminable via lipreading alone in either initial or final position. The left half of Figure 5-2 (top and bottom) shows that, in CVC-initial position, /s/ and /t/ are discriminated at near-chance levels in the V-alone condition, which is comparable to discrimination performance for the six voicing contrast pairs. The right half of Figure 5-2 reveals that, in CVC-final position, V-alone discrimination of the /s/-/t/ manner contrast is actually poorest among the seven consonant pairs.

With the addition of tactile cuing, consonant discrimination performance improved substantially across subjects. Discrimination of some pairs was more consistently and robustly enhanced in the V+T condition, as is reflected in Figure 5-2. Discrimination of the /s/-/t/ manner contrast showed the most benefit from tactual supplementation of lipreading across subjects, in both initial and final CVC contexts. All subjects successfully exploited the tactual manner cues, with the exception of subject HI4, whose elevated index finger detection thresholds are noted above. Both /s/ and /t/ are characterized by high frequency noise associated with an alveolar constriction. With most spectral energy falling within the high frequency speech band (i.e., above 3000 Hz), both /s/ and /t/ give rise to distinct patterns of tactual stimulation confined mainly to the index finger tactual channel. In both initial and final CVC contexts, the two tactual patterns are distinguished by their amplitude envelopes. The percept at the index finger corresponding to an /s/ has subtle onset and offset, often blending perceptually with the adjacent vowel. By contrast, the burst of the /t/ produces a sudden onset with a discrete percept. In CVC-final context, the /t/ burst is quite noticeably preceded by a cross-channel lull (corresponding to full vocal tract constriction) and is therefore quite distinct from the preceding vowel. In isolation, the /s/-/t/ discrimination task is likely facilitated by the relatively lax requirement of attending to the overall amplitude envelope at a single site, rather than a cross-channel pattern progression.

Tactual cuing substantially enables discrimination of all six voicing contrast pairs, in both initial and final syllabic contexts. However, the ability to exploit tactile information varied among subjects. In CVC-initial position, tactual cues distinguishing the consonant pairs /t/-/d/ and /p/-/b/ were most readily discerned by subjects in the first procedural repetitions. As such, these pairs might be considered the most easily discriminable among the six voicing pairs. By contrast, the tactual cues distinguishing the consonant pairs /f/-/v/ and /tʃ/-/dʒ/ were less readily discriminated in initial position, and several subjects

acquired the tactual distinction only during their second procedural repetitions (i.e., after extensive practice).

The relative discriminability of the various component pairs may best be understood through inspection of the corresponding vibrotactile waveforms, samples of which are depicted in Figure 4-6. For example, in the case of the syllable-initial /tʃ/-/dʒ/ distinction (Figure 4-6, bottom-right), one may observe that the initial /dʒ/ waveforms exhibit a substantial cross-channel onset-asynchrony, with the I-channel onset preceding the M-channel onset by nearly 100 ms. Such a vibrotactile signal, presented in the absence of visual cues, might reasonably be interpreted as an unvoiced consonant --- e.g., note in Figure 4-6 that the initial /dʒ/ and initial /t/ vibrotactile waveforms differ only subtly. Successful recognition of the voiced initial /dʒ/ requires integration of tactual cues with visual place and/or manner cues, as well as familiarity with the vibrotactile signals corresponding to both the voiced and unvoiced palato-alveolar consonants (/dʒ/ and /tʃ/, respectively). Only with experience does one come to differentiate the I-channel burst of /dʒ/, which merges seamlessly the subsequent vowel, the burst of /t/, which is temporally distinct from the vowel, and the burst of /tʃ/, which has a similarly sharp onset but substantially longer duration than either /t/ or /dʒ/ (as well as being temporally distinct from the vowel).

Figures 5-3 and 5-4 reveal several distinct patterns of performance among the eight participants in this study. The most common pattern is that exhibited by subjects NH1, NH2, NH3, HI1, and HI3, all of whom clearly derive substantial benefit from tactual cuing for discrimination of CVC-initial and final consonants. In each case, tactual benefit is evident from the first procedural repetition and persists (or increases further) in cases where a second repetition was performed.

Subject HI2 shows significant tactual benefit for initial consonant discrimination, but not for final consonant discrimination. This performance is consistent across both repetitions. Subject HI4 shows no significant tactual benefit at all for pairwise consonant discrimination in either initial or final CVC context. Subject HI5 performs initial and final consonant discrimination at chance levels under both V-alone and V+T sensory conditions in her first repetition. In her second repetition, however, she exhibits significant tactual benefit in both initial and final CVC contexts. The substantial deviation of these subjects' performance from that of the other five participants is considered further below.

Bearing in mind that our ultimate objective is to prepare tactile-aid users for reception of running speech, requiring identification of both initial and final consonants in the course of a single experimental trial was considered preferable to training 12-consonant identification in initial and final positions separately. This approach encouraged subjects to attend and integrate speech at the syllable level, rather than focusing on individual phonetic elements.

In general, individual performance in the CVC identification task under the V+T condition reflected V+T pairwise discrimination performance. Comparing relative performance in pairwise initial (Figures 5-3, top) and final (Figures 5-4, top) consonant discrimination with the corresponding CVC identification scores (Figure 5-5), we find that the four top performers in initial consonant pairwise discrimination (NH1, NH2, NH3, and HI3) all demonstrate roughly to 80% correct initial consonant identification. These same individuals, along with subject HI1, are the five top performers in final consonant pairwise discrimination, and they each demonstrate roughly 90% correct final consonant identification.

Although HI1 does not perform as well as the other four subjects in initial consonant reception, she shows a marked increase in average sensitivity for initial consonant contrasts under the V+T condition between the first and second procedural repetitions, in which her d' values fall close to 1 (roughly 70% correct) and 2 (roughly 85% correct), respectively (as observed in the top panel of Figure 5-3). Similarly, Figure 5-5 (top) indicates that, under the V+T condition, HI1 correctly identifies approximately 70% of initial consonants in the 12-alternative evaluation of her second repetition, which represents an improvement of roughly 10% from her first repetition score.

Subject HI2, one of the two individuals with elevated tactual detection thresholds, demonstrated moderate tactual benefit for initial consonant pairwise discrimination ($d' \approx 1.5$) and no significant benefit for final consonant pairwise discrimination. For both CVC contexts, her average sensitivity scores were nearly unchanged between her first and second repetitions. Not surprisingly, her performance in the 12-consonant identification task was also quite similar in the first and second repetition. She demonstrated a tactual benefit of roughly 10% for initial consonant identification, and 7-8% tactual benefit for final consonant identification. In all cases, the improvement in her identification performance in the V+T relative to the V-alone condition was found to be statistically significant. Interestingly, although HI2 showed no significant difference in sensitivity when averaged across final consonant pairs, her performance on the individual pairs (data not shown) indicates that, by the second repetition, she demonstrated near-perfect V+T discrimination of the syllable-final /t/-/s/ contrast and roughly 85% correct V+T discrimination of the syllable-final /tʃ/-/dʒ/ contrast, both of which exceeded her V-alone

scores (about 50% and 70% correct, respectively). However, her discrimination performance for each of the other five final consonant pairs was either comparable under the two sensory conditions, or else it was better under the V-alone condition (suggesting that focusing on the tactual signal may actually have impaired her speechreading performance in these cases).

This peculiar performance exhibited by subject HI2 is, in fact, quite easily understood in light of her substantially elevated 200 Hz detection threshold at the left middle finger. Under the current speech transduction scheme, such a deficit in tactual sensitivity would be expected to impair one's perception of voicing-induced vibrations deriving from the low-frequency portion of the acoustic spectrum (i.e., the voicing fundamental and lower harmonics). However, looking at the vibrotactile waveforms of Figure 4-6, one may observe that, in CVC-final context, both the /t-/s/ and /tʃ-/dʒ/ contrasts can potentially be discriminated based on the burst duration observed on the I-channel (index finger) signal alone. Such is not the case for the other five consonant pairs, as the I-channel burst amplitude is not an invariant indicator of consonant voicing.

Subject HI4, whose left index finger detection thresholds were elevated at all three frequencies tested, demonstrated no significant tactual benefit for discrimination of either initial or final consonant contrast pairs. In her first repetition (in which she trained non-vocally), she also exhibited no tactual benefit in the 12-consonant identification task. Somewhat surprisingly, however, her initial and final consonant identification performance in the second repetition (in which she trained vocally) both improved significantly ($p < .01$) in the V+T condition relative to V-alone. The tactual benefit observed was close to 20% increased accuracy for initial consonant identification and about 5% for final consonant identification. Figures 5-7 and 5-8 suggest that, in the second repetition, HI4 showed improved reception of both voicing and articulatory manner under the V+T condition. This is not readily explained in terms of her pairwise discrimination performance.

Finally, subject HI5 exhibited no significant tactual benefit for either consonant discrimination or identification during her first procedural repetition, in which she trained non-vocally. Following vocal training in her second repetition, she demonstrated significant tactual benefit for both pairwise discrimination and 12-consonant identification, in both syllable-initial and final positions. Her discrimination performance for individual consonant pairs (not shown) indicate that, during the first repetition, she derived tactual benefit only for the /t-/d/ contrast in CVC-final context ($d' \approx 1$ in the V+T condition). However, in her second repetition, she exhibited tactual benefit for the CVC-initial consonant

pairs /t-/s/, /t-/d/, /p-/b/, and /k-/g/, as well as for the CVC-final consonant pairs /t-/s/, /t-/d/, /k-/g/, and /f-/v/.

Of the five profoundly deaf participants, HI5 was the only one who was neither a cochlear implantee nor a hearing aid wearer. Her speech was often unintelligible (e.g., she frequently substituted /s/ for other alveolar consonants). Among all of the participants, HI5 was one of two individuals (the other being HI1) to indicate unambiguously that she found the vocal training strategy more useful and engaging.

5.4.5 Performance Variability among Subjects

Participants in this study exhibited substantial variability in their ability to discriminate even the most tactually distinct consonant feature-contrast pairs. Such performance variability at the pairwise discrimination level was not observed by Yuan et al. (2005b).

Importantly, the discrepancy is not readily explained by the inclusion of hearing-impaired participants in the present study. Two of the five hearing-impaired subjects acquired the tactually aided consonant discrimination task in a manner consistent with the performance of normal-hearing subjects observed by Yuan and colleagues. Moreover, it should be noted that, apart from the three normal-hearing subjects who completed at least one full repetition in the current study, four additional normal-hearing individuals, two males and two females ranging in age from 20-32 years, started the study but withdrew prior to completing the initial pairwise consonant discrimination phase. Although they did not undergo formal evaluation, all participated in at least three experimental sessions, during which two of the four individuals showed little deviation from chance performance. (Notably, one the two chance-performers was observed on multiple occasions to have fallen asleep in an upright seated position, in the middle of a training run --- in each case, he was left undisturbed until, startling himself awake, he resumed near-chance performance.)

Of the three hearing-impaired subjects who completed two procedural repetitions in the current study and yet showed somewhat limited tactually-aided consonant discrimination and identification performance, two of them (HI2 and HI4) exhibited elevated tactual detection thresholds of a nature that could reasonably be expected to have compromised their tactual speech reception. As noted above, the third of these individuals, subject HI5, was the only one among the five profoundly deaf participants who was neither a cochlear implantee nor a hearing-aid user. HI5 also exhibited the poorest speech communication skills, including both speechreading and intelligibility, among the five hearing-impaired participants. She

was the only subject with whom the experimenter was regularly required to supplement verbal communication with handwritten clarification.

With that being said, the results of this study leave no doubt that the tactual transduction scheme under evaluation can effectively support consonant discrimination and identification in CVC context. The consonant discrimination performance of subjects NH1, NH2, NH3, HI1, and HI3 during their first procedural repetitions was comparable to that observed by Yuan et al. (2005b), whose results are reproduced in Figure 4-1 of the previous chapter. Note that Yuan et al. used a two-interval 2AFC testing procedure, and their reported mean d' values (rightmost bars in Figure 4-1) must be divided by a factor of $\sqrt{2}$ (roughly 1.41) for purposes of comparison with the corresponding one-interval 2AFC d' values (Macmillan and Creelman, 1990) of subjects NH1, NH2, NH3, HI1, and HI3, depicted in the top panel of Figure 5-3. Specifically, Yuan et al. report a mean d' of about 2.5 in the "L+T" (lipreading + tactual) condition, averaged across subjects and consonant pairs, which divided by $\sqrt{2}$ yields an equivalent one-interval d' value of approximately 1.8. By comparison, in their first procedural repetitions, Figure 5-3 indicates that d' values averaged across consonant pairs for NH1, NH2, NH3, HI1, and HI3 were roughly 1.6, 2.3, 2.8, 1.1, and 2.3, respectively. Subjects NH1, NH2, and HI1 each performed second repetitions, in which their mean d' scores increased to roughly 2.7, 2.5, and 2.0 respectively.

Group-based comparison of normal-hearing and hearing-impaired subjects is confounded by several factors. Most notably, subjects HI2 and HI4 exhibited elevated tactual detection thresholds that one might reasonably expect to impair tactual speech discrimination performance in this study. On these grounds, data from HI2 and HI4 would best be excluded from any hearing-status based group comparison. Moreover, the five profoundly-deaf participants are a rather heterogeneous group, differing in basic etiology of hearing loss and mode of communication during childhood, as well as in their current communication skills and sensory aid usage. Thus, a direct comparison of tactual speech performance among normal-hearing and hearing-impaired participants is not feasible. However, it is not at all clear that restriction of hearing-impaired participants on the basis of etiology, communication mode, and sensory aid usage would have provided results of more practical import. The fact of the matter is that individuals with profound hearing-impairment are a very heterogeneous population. It was this simple reality that, in the course of recruiting hearing-impaired subjects to participate in this study, led us to accept such a diverse group.

Beyond the issues of elevated tactual detection thresholds and limited speech communication skills, certain experimental features of the current study might further account for any remaining unexplained

discrepancy with the results of Yuan and colleagues. A major difference in the training protocols of the two studies (aside from the sensorimotor element) is the omission of the tactual-alone (T-alone) sensory condition from the current study. The decision to omit the T-alone condition was taken to reduce experimental time by one-third, leaving only the V-alone and V+T conditions. As a result of this decision, subjects were never required to make articulatory feature distinction based solely on tactual inputs. Considering, in particular, that pairwise discrimination of CVC-initial consonants depended almost exclusively upon tactual cues in this study, presentation of speakers' faces (i.e., the visual component) under the V+T condition served primarily as a distraction. During this phase of training, subjects regularly expressed dismay at their chance performance in V-alone trials, suggesting that they maintained the belief that they were failing to recognize visual cues. By contrast, Yuan et al. (2005b) trained subjects under the tactual-alone condition, in addition to the visual-alone and visual+tactile conditions. Thus, subjects could not help but recognize the fact that they could successfully discriminate consonant pairs based solely on tactual cues, and it would be only natural to exploit these cues under the visual+tactile condition as well. In the current study, however, it is possible that omission of the T-alone condition rendered consonant discrimination training substantially less effective than it might otherwise have been. Future trainings should incorporate the T-alone condition and may thereby allow evaluation of this hypothesis.

Another aspect of the present study that could have adversely affected the outcome is the inclusion of the thumb channel (corresponding to the mid-frequency acoustic band) from the outset of training. It is conceivable that subjects' acquisition of the tactual cues enabling pairwise consonant discrimination would have proceeded with less confusion had the thumb channel remained inactive during the first few experimental sessions. The thumb channel carries potentially useful information that is perceptually discriminable, at least when experienced in isolation. However, for the six consonant voicing contrasts, the primary tactual cue relied most heavily on differences in the timing of signals delivered to the index and middle fingers. In the early stages of training, the presence of a third channel may have distracted the subjects, obfuscating the EOA and EOFA tactual cues to some extent. Toward the end of their participation, subjects NH1, NH2, and HI1 indicated that their perceptual strategies ultimately involved neglecting the thumb channel signal in order to focus attention on the primary voicing and manner cues delivered to the index and middle fingers.

5.4.6 Consideration of Sensorimotor Training Strategy

Among the objectives of this study was the evaluation of a novel approach to training tactile speech reception. The rationale for a perceptual training protocol that integrates sensorimotor feedback is to facilitate the learning of tactual speech cues by associating them with corresponding articulatory acts. The underlying hypothesis is that, when introducing speech through a novel sensory modality, the process of augmenting the CNS representation of speech would benefit from the conditional association of the novel sensory patterns with the corresponding articulatory gestures and proprioceptive activation patterns, already intrinsic to cortical speech representation. The results of this study indicate that the benefit derived from augmenting training by incorporating tactual feedback of one's own vocalizations appears to differ among subjects.

A group-based analysis (multi-factor ANOVA) considering the effects of sensory condition, training strategy, and procedural repetition on the performance of all subjects revealed significant effects for sensory condition only. However, the variability in subject performance at the most elementary level of the experiment (pairwise discrimination) was not conducive to objective evaluation of a novel perceptual training approach. Perhaps closer to the heart of the issue is the possibility that inclusion of the sensorimotor element in the consonant discrimination training phase was not equally necessary for all subjects. In reviewing Figures 5-3 and 5-4, consider that subjects NH1 and NH3 benefitted substantially from tactual cuing in their first repetitions of the consonant discrimination evaluation, despite their both having undergone non-vocal training. Since no special training protocol is required to achieve substantial V+T benefit in a single training repetition, there is then no basis for linking the substantial V+T benefit demonstrated by NH2, HI1, and HI3 to their having undergone vocal training in the first repetition.

A two-way repeated-measures ANOVA revealed a significant main effect of training strategy on initial consonant identification performance ($p = .036$), qualified by a significant interaction between training strategy and training order ($p = .016$). The effect of training strategy indicates a significant difference in the efficacy of the NV and VF training strategies. Note that this analysis does not explicitly differentiate between the first repetition, in which subjects' performance reflects a single training experience, and the second repetition, in which performance reflects the cumulative effect of both training strategies. Subjects in the NV-VF group (HI4 and HI5), who received NV training first, showed tactile speech-benefit *only* after their second (VF) training. By contrast, subjects in the VF-NV group (HI1 and HI2), who received VF training first, demonstrated tactile speech-benefit after their first training and again after their second (NV) training. The interaction between training strategy and training order indicates that the

two groups appear to demonstrate a significant difference in relative benefits from the two training strategies; the VF-NV group perform well after both trainings, while the NV-VF group perform well only after the second training. Since performance after the second training reflects the cumulative effects of both trainings, the significant interaction between training strategy and training order indicates a difference in the amount of performance benefit carried over from the first training to the second training. As such, both the main effect of training strategy and the interaction effect indicate the substantial tactile speech-benefits acquired under the VF training strategy relative to the NV training strategy.

To summarize, the results of this study indicate that, for the three normal-hearing subjects (NH1, NH2, NH3) and the one profoundly deaf subject (HI3), who learned quickly to exploit tactual speech cues, the sensorimotor training strategy implemented in this study may have been unnecessary. However, for the four hearing-impaired subjects who learned to exploit the tactual cues more gradually, the sensorimotor training strategy appeared to benefit learning of tactually-supplemented speechreading of CVC monosyllables.

5.5 Summary

Tactual Psychophysics:

- Tactual detection thresholds at 10 Hz, 50 Hz, and 200 Hz, were measured at the left thumb, index finger, and middle finger of each subject. Excluding those of HI2 and HI4, all other subjects' detection thresholds for each finger-frequency combination were less than one s.d. above the corresponding group mean threshold. Overall, detection thresholds are consistent with measurements reported in Chapter 3 of this dissertation and in previous studies. [Figure 5-1; Table 5-3]
- Subject HI2 exhibited elevated thresholds for all three fingers at 200 Hz. Subject HI4 exhibited elevated thresholds at the left index finger for all three frequencies tested. The threshold elevation patterns of both HI2 and HI4 may have prevented them from effectively exploiting tactual speech cues in this study.
- The average value of tactual onset-order thresholds across subjects is 175.7 ms, with a standard deviation of 96.5 ms. The lowest observed onset-order threshold was 52.8 ms (subject NH3), and the highest observed was 294.1 ms (subject HI5). [Table 5-4]

Pairwise Consonant Discrimination:

- Under the V-alone condition, mean syllable-initial consonant discrimination performance across subjects is close to chance for all seven consonant pairs, whereas mean syllable-final consonant discrimination performance across subjects is close to chance only for the manner contrast pair, /t/-s/. Unlike initial consonant voicing, final consonant voicing often accompanied by vowel lengthening, which may be visually detectable as a lengthening of the utterance. [Figure 5-2]
- Mean initial consonant discrimination across subjects under the V+T condition significantly exceeds V-alone performance ($p < .01$) for six of the seven consonant pairs (all but /f/-v/) after the subjects' first procedural repetitions. Following the subjects' final repetitions, the difference is significant ($p < .01$) for all seven pairs. [Figure 5-2]
- Mean final consonant discrimination across subjects under the V+T condition significantly exceeds V-alone performance for the consonant pairs /t/-s/ ($p < .01$) and /k/-g/ ($p < .05$) after the subjects' first procedural repetitions. Following the subjects' final repetitions, the difference is significant for the consonant pairs /t/-s/, /t/-d/, and /k/-g/ ($p < .01$), as well as /f/-v/ and /dʒ/-tʃ/ ($p < .05$). [Figure 5-2]

- For discrimination of syllable-initial consonant pairs, each individual subject averages close to $d' = 0$ under the V-alone condition. By contrast, under the V+T condition, five out of eight subjects demonstrate performance exceeding $d' = 2$ (roughly 85% correct response rate), and an additional two subjects exceed the $d' = 1$ criterion. [Figure 5-3]
- For discrimination of syllable-final consonant pairs, seven out of eight subjects demonstrate mean sensitivity scores greater than $d'=1$ under the V-alone condition (in all repetitions performed). Under the V+T sensory condition, pairwise discriminability of syllable-final consonants improved significantly ($p < .05$) for six out of eight subjects. Only subjects HI2 and HI4 (the two individuals exhibiting elevated tactile detection thresholds) performed comparably under the V+T and V-alone conditions in both their first and second procedural repetitions. [Figure 5-4]

Twelve-Consonant Identification:

- Initial consonant identification performance falls roughly in the range of 40-50% accuracy for all subjects in the V-alone condition and 50-80% accuracy in the V+T condition. T-tests show that V+T performance is significantly better than V-alone performance ($p < .01$) for all subjects in all repetitions, with the exception of subjects HI4 and HI5 in the first repetitions. [Figure 5-5, top]
- Final consonant identification performance in the V-alone condition ranges from 60-80% for all subjects other than HI5, who performs at about 40% accuracy. Performance in the V+T condition approaches or exceeds 90% accuracy (and significantly exceeds V-alone performance at $p < .01$) for five out of eight subjects (NH1, NH2, NH3, HI1, HI3). Of the remaining three subjects (HI2, HI4, HI5), only HI2 shows significant benefit in V+T relative to V-alone ($p < .05$) in the first repetition. However, in the second repetition, all three subjects show significant benefit in V+T relative to V-alone, at a level of $p < .01$. [Figure 5-5, bottom]
- Three-way ANOVAs for performance effects of sensory condition (V-alone, V+T), training strategy (NV, VF), and experimental repetition (first, second) show significant effects only for sensory condition, in both CVC-initial ($p < .00001$) and CVC-final ($p = .02$) consonant identification.
- Two-way, repeated-measures ANOVAs were conducted to test for effects of training strategy and training order on tactual speech-benefit to CVC-initial and CVC-final consonant identification performance among the four hearing-impaired subjects who completed two full procedural repetitions. For initial consonant identification performance, a significant main effect of training strategy ($p = .036$) was observed, qualified by a significant interaction between training strategy and training order ($p = .016$). For final consonant identification performance, no significant effects were observed for training strategy, training order, or interaction between them.

- Initial consonant voicing reception averaged less than 2% IT in the V-alone condition and about 31% IT in the V+T condition in subjects' final repetitions. Final consonant voicing reception averaged about 36% IT in the V-alone condition, and 68% IT in the V+T condition. [Figure 5-9, left]
- Initial consonant manner reception averaged roughly 48% IT in the V-alone condition and 72% IT in the V+T condition in subjects' final repetitions. Final consonant manner reception averaged about 69% IT in the V-alone condition and 85% IT in the V+T condition. [Figure 5-9, center]
- Initial consonant place reception averaged about 78% IT in the V-alone condition and 81% IT in the V+T condition in subjects' final repetitions. Final consonant place reception averaged about 81% IT in the V-alone condition and 85% IT in the V+T condition. [Figure 5-9, right]

Correlation Analyses:

- Tactual onset-order thresholds are inversely correlated with syllable-initial consonant discrimination in the V+T condition. The magnitude of this correlation is larger for subjects' first repetitions (correlation coefficient $r = -0.976$, significant at $p < .01$) than for subjects' final repetitions (correlation coefficient $r = -0.774$, significant at $p < .05$), suggesting that the correlation between temporal resolution and V+T initial consonant discrimination diminishes with training. [Table 5-6]
- Negative correlation coefficients between subjects' observed tactual speech-benefit and their tactual detection thresholds suggest a correlation between higher detection thresholds (lower vibrotactile sensitivity) and reduced tactual benefit to speech reception. At the left index finger and thumb, significant negative correlations are observed between 200 Hz detection thresholds and tactual benefit to identification of CVC-initial and -final consonants. At the left middle finger, significant negative correlations are between 10 Hz, 50 Hz, and 200 Hz detection thresholds and tactual benefit to identification of CVC-initial and -final consonants, but for 200 Hz, these correlations are significant ($p < .05$) only for subjects' final repetitions. [Table 5-7]

5.6 Concluding Remarks

The extreme variability of tactual temporal order thresholds among both normal-hearing and hearing-impaired participants in the current study has fundamental implications for the next phase of this project. The apparent discrepancy between the two sets of onset-order measurements reported in Chapters 3 and 5 of this dissertation suggests that ability to discriminate temporal order across tactual channels may vary substantially, not only between individuals, but based on behavioral context and specific task requirements. In the interest of providing a single tactual speech strategy that benefits the greatest number of users and demands the minimum possible training commitment, the present encoding strategy might best be augmented so as to render temporal patterns more readily discernable as positional information, rather than fast spatiotemporal patterns.

On the other hand, one must not overlook the fact that individuals such as subjects NH1 and HI1, whose respective tactual onset-order thresholds measured 174 and 220 ms, were nonetheless able to distinguish voiced from unvoiced consonants quite effectively in the present study. In fact, NH1 performed comparably in all tactual speech tasks to subjects NH2 and NH3, whose respective tactual onset-order thresholds measured 55 and 53 ms. Of course, it is important to recognize that the stimulus parameters used to examine onset-order discrimination offer at best a much simplified model for the reception of tactual EOA cues. The chosen model focuses exclusively on cross-channel temporal ordering, which is but one aspect of the speech transduction strategy implemented in this study to facilitate consonant discrimination and identification. The relative contributions of cross-channel spatiotemporal cues and within-channel spectral cues to tactile speech reception were not examined directly. It is quite possible that cross-channel temporal order discrimination was facilitated by the redundancy of coding across multiple somatosensory submodalities (e.g., kinesthetic, flutter, vibratory/Pacianian). Thus, despite being unable to discriminate the onset-order of two 200 Hz signals initiated 100 ms apart at the index and middle fingers, subject NH1 may be able to discriminate much finer temporal asynchronies when at least one of the stimuli has a transient, large-amplitude component, as is commonly observed on the I-channel in Figure 4-6. The relative contributions of large-amplitude displacements in the kinesthetic domain and small-amplitude vibrations activating Pacinian receptors at each tactual locus should be examined prior to further modification of the current speech transduction scheme.

The chance inclusion in this study of individuals with elevated tactual detection thresholds underscores the need for rudimentary sensitivity testing prior to fitting tactual prosthesis to a particular body site. If a single wearable configuration is desired, whether actuators are to contact the fingers of one hand (as in the

present work) or perhaps migrated to a less obtrusive body location, it will likely be necessary to adjust the signal gain on each channel in accordance an individual's site-specific detection thresholds. In a sense, gain adjustments of this sort might be regarded as analogous (in a rudimentary way) to the fitting of a hearing-aid, based on an individual's specific pattern of hearing loss. However, tactual masking patterns differ substantially from auditory masking patterns, and the latter are far better characterized. Adjusting the relative gains of several vibrotactile channels in order to fit an individual's local sensitivity function might, in turn, lead to substantial variability in cross-channel interference between individuals. A more practical option might therefore involve fitting each individual with one of several possible stimulator configurations, depending on one's psychophysical measurements and personal preferences. Thus, in the interest of accommodating as many beneficiaries as possible, future research should include evaluation of the tactual speech transduction scheme presented here, adapted for placement of transducers at alternative body sites in conveniently wearable configurations.

Although inclusion of the middle frequency (thumb) channel was intended to provide tactual cues for gross formant structure and articulatory place (alveolar vs. velar in particular), there is little evidence that this channel benefitted subjects' perception of consonant place in this study. Rather, the presence of this third, mid-frequency channel may have hampered tactual speech acquisition for some individuals, particularly in the early stages of perceptual training, by obscuring cues for consonant voicing and articulatory manner present mainly on the index and middle finger channels. In future work, this channel will likely be omitted from the tactual display until the user can reliably discriminate consonant voicing and manner. Thereafter, the benefit imparted by adding a third, mid-frequency channel can be more effectively determined.

Overall, the current study demonstrates the ability of profoundly deaf and normal-hearing individuals to exploit tactual cues derived from an acoustic speech signal to substantially facilitate speechreading of monosyllabic utterances. Specifically, tactual cuing is observed to support identification of consonants in both syllable-initial and syllable-final contexts by conveying articulatory manner and voicing information that is not otherwise available through lipreading alone. Future work will focus on two parallel objectives: (1) development of a low-cost hardware platform capable of extending the benefits of tactile speech aids to hearing-impaired individuals in the developing world, and (2) assessment of an integrated strategy for training tactually-aided speech production and visual speech recognition.

In addition to conveying consonant voicing and manner information to supplement visual-facial cues, tactual cuing (as represented in Figure 4-6) can preclude a vast number of perceptual confusions when

two consonants differ in both articulatory place and manner, thereby supporting speechreading even when visual information is heavily degraded. The tactile display strategy described in Chapter 4 could thus offer distinct benefits to hearing-impaired individuals who also have some degree of visual impairment. For example, tactual cuing could significantly postpone the loss of speech reception via lipreading for deaf individuals with progressive vision loss, commonly experienced in Usher's Syndrome (Vernon, 1969), which one would expect to have a substantially positive impact on such individuals' quality of life. Although not addressed directly by this thesis, the development of audio-tactile transduction technologies that meet the particular sensory needs of deafblind individuals represents a distinct application of this line of research that will ultimately require attention in its own right.

References

- Alcorn, S. K. (1932). The Tadoma Method. *The Volta Review*, May, 195-198.
- Assmann PF. (1999) Fundamental Frequency and the Intelligibility of Competing Voices. 14th International Congress of Phonetic Sciences, San Francisco, 1-7 August 1999.
- Aytekin M, Moss CF, Simon JZ. (2008). A Sensorimotor Approach to Sound Localization. *Neural Computation* 20, 603–635.
- Bassim, M. K., Buss, E., Clark, M. S., Kolln, K. A., Pillsbury, C. H., Pillsbury, H. C., III, Buchman, C. A. (2005). "MED-EL Combi40 cochlear implantation in adults," *Laryngoscope* 115, 1568-1573.
- Békésy G. v. (1957). Sensations on the Skin Similar to Directional Hearing, Beats, and Harmonics of the Ear. *J Acoust Soc Am.* 29(4): 489-501.
- Bernstein LE, Eberhardt SP, Demorest ME. (1989) Single-channel vibrotactile supplements to visual perception of intonation and stress. *J Acoust Soc Am.* 85(1):397-405.
- Bernstein LE, Demorest ME, Coulter DC, O'Connell MP. (1991). Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *J Acoust Soc Am.* 90(6): 2971-2984.
- Bernstein LE, Demorest ME, Eberhardt SP. (1994). A computational approach to analyzing sentential speech perception: Phoneme-to-phoneme stimulus–response alignment. *J Acoust Soc Am.* 95(6): 3617-3622.
- Bird J, Darwin CJ. (1998). Effects of a difference in fundamental frequency in separating two sentences. In AR Palmer, A Rees, AQ Summerfield, R Meddis (Eds.), *Psychophysical and physiological advances in hearing*. London: Whurr.
- Blessner BA. (1969) Ph.D. Thesis. Perception of spectrally rotated speech. Department of Electrical Engineering, Massachusetts Institute of Technology.
- Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky C. M. (1988). "Four channels mediate the mechanical aspects of touch," *J. Acoust. Soc. Am.* 84, 1680-1694.
- Boothroyd A, Hnath-Chisolm T. (1988) Spatial, tactile presentation of voice fundamental frequency as a supplement to lipreading: results of extended training with a single subject. *J Rehabil Res Dev.* 25(3):51-6.
- Boothroyd A, Hnath-Chisolm T, Hanin L, Kishon-Rabin L. (1988) Voice fundamental frequency as an auditory supplement to the speechreading of sentences. *Ear Hear.* 9(6):306-12.
- Bornstein H. (1990). *Manual communication: Implications for education*. Gallaudet University Press, Washington, D.C.

- Bradlow AR, Pisoni DB, Yamada RA, Tohkura Y. (1997) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J Acoust Soc Am.* 101(4), 2299.
- Braida LBD. (1965) Bilocal Cutaneous Unmasking. Masters Thesis, MIT.
- Bratakos MS, Reed CM, Delhorne LA, Denesvich G. (2001) A Single-Band Envelope Cue as a Supplement to Speechreading of Segmentals: A Comparison of Auditory versus Tactual Presentation. *Ear and Hearing*, Volume 22(3), 225-235.
- Breeuwer M, Plomp R. (1984) Speechreading supplemented with frequency-selective amplitude-envelope information. *J Acoust Soc Am.* 76, 686-691.
- Brokx JPL, Nooteboom SG. (1982). Intonation and the perception of simultaneous voices. *Journal of Phonetics*, 10, 23-26.
- Brooks PL, Frost BJ, Mason JL, Gibson DM. (1986). Continuing evaluation of the Queen's University tactile vocoder. I: Identification of open set words. *J Rehabil Res Dev.* 23(1):119-28.
- Brooks PL, Frost BJ, Mason JL, Gibson DM. (1986). Continuing evaluation of the Queen's University tactile vocoder II: Identification of open set sentences and tracking narrative. *J Rehabil Res Dev.* 23(1):129-138.
- Browman CP, Goldstein LM. (1986) Towards an articulatory phonology. In C. Ewen and J. Anderson (eds.) *Phonology Yearbook 3*. Cambridge: Cambridge University Press, pp. 219-252.
- Browman CP, Goldstein L. (1992). Articulatory phonology: An overview. *Phonetica.* 49, 155-180.
- Brown, L. N., and Sainsbury, R. S. (2000). "Hemispheric equivalence and age-related differences in judgments of simultaneity to somatosensory stimuli," *J. Clin. Exp. Neuropsychol.* 22(5), 587-98.
- Buonomano D, Merzenich MM. (1998). CORTICAL PLASTICITY: From Synapses to Maps. *Annual Review of Neuroscience.* Vol. 21: 149-186.
- Chakravarty, A. (1968). "Influence of tactual sensitivity on tactual localization, particularly of deaf children," *J. Gen. Psychol.* 78, 219-221.
- Chomsky C. (1986). Analytic study of the Tadoma method: language abilities of three deaf-blind subjects. *J Speech Hear Res* 29(3): 332-47.
- Clark, J. G. (1981). "Uses and abuses of hearing loss classification," *ASHA*, 23, 493-500.
- Clements MA, Braida LD, Durlach NI. (1982) Tactile communication of speech: II. Comparison of two spectral displays in a vowel discrimination task. *J Acoust Soc Am.* Oct;72(4):1131-5.
- Cooper W. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex. Pub. Corp.
- Cornett RO. (1988). Cued speech, manual complement to lipreading, for visual reception of spoken language: Principles, practice and prospects for automation. *Acta Otorhinolaryngol Belg.* 42(3):375-84.

- Craig, J.C. (1972). Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics*, 11(2), 150-152.
- Craig, J. C., Baihua, X. (1990). "Temporal order and tactile patterns," *Percept. Psychophys.* 47, 22-34.
- Craig, J. C., Evans, P. M. (1987). "Vibrotactile masking and the persistence of tactual features," *Percept. Psychophys.* 42, 309-317.
- Craig, J. C., Busey, T. A. (2003). "The effect of motion on tactile and visual temporal order judgments," *Percept. Psychophys.* 65, 81-94.
- Craig, J.C., Rollman G.B. (1999). "Somesthesia," *Annu. Rev. Psychol.* 50:305-331.
- Cranney, J., and Ashton, R. A. (1982). "Tactile spatial ability: Lateralized performance of deaf and hearing age groups," *J. Exp. Child Psychol.* 34, 123-134.
- Denes P. (1955) "Effect of Duration on the Perception of Voicing," *J. Acoust. Soc. Am.* Volume 27, Issue 4, pp. 761-764.
- Dillon, H. (2001). *Hearing Aids*. New York: Thieme; Sydney: Boomerang Press.
- Dodd B. (1977). "The role of vision in the perception of speech," *Perception* 6, 31-40.
- Dodd B, Plant G, Gregory M. (1989) Teaching lipreading: The efficacy of lessons on video. *Br. J. Audiol.* 23, 229-238.
- Durlach, N. I. (1968). "A decision model for psychophysics," *Communication Biophysics Group, Research Laboratory of Electronics, MIT, MA.*
- Eberhardt, S. P., Bernstein, L. E., Barac-Cikoja, D., Coulter, D. C., and Jordan, J. (1994). "Inducing dynamic haptic perception by the hand: System description and some results," *Proceedings of the ASME Dynamic Systems and Control Division, New York, Vol.1, pp. 345-351.*
- Eberhardt SP, Bernstein LE, Demorest ME, Goldstein MH Jr. (1990) Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency. *J Acoust Soc Am.* 88(3):1274-85.
- Eimas PD, Corbit JD. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- Enerstvedt RT. (1999). *Theodor - Lexicon of Impairments*. Department of Sociology and Human Geography. Harriet Holters hus, Postboks 1096 Blindern, 0317 Oslo. [formerly available at <http://home.wit.no/skarpreklame/Kunder/Bedrifter/Theodor/>]
- Erber NP. (1974). Visual perception of speech by deaf children: Recent developments and continuing needs. *Journal of Speech and Hearing Disorders*, 39:178-185.
- Friel-Patti S, Roeser RJ. (1983). Evaluating changes in the communication skills of deaf children using vibrotactile stimulation. *Ear Hear.* 4(1):31-40.

- Fowler CA. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Galantucci B, Fowler CA, Turvey MT. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*. 13(3), 361-377.
- Galvin KL, Mavrias G, Moore A, Cowan RS, Blamey PJ, Clark GM. (1999) A comparison of Tactaid II+ and Tactaid 7 use by adults with a profound hearing impairment. *Ear and hearing*. 20(6):471-82.
- Gault, R. H. (1924). "Progress in experiments on tactual interpretation of oral speech," *J. Abnorm. Soc. Psychol.* 19, 155-159.
- Gault, R. H. (1926). "Touch as a substitute for hearing in the interpretation and control of speech," *Arch. Otolaryngol.* 3, 121-135.
- Gault, R. H. (1930): "On the effect of simultaneous tactual-visual stimulation in relation to the interpretation of speech," *J. Abnorm. Social Psychol.* 24, 498-517.
- Gault, R. H., and Crane, G. W. (1928). "Tactual patterns from certain vowel qualities instrumentally communicated from a speaker to a subject's fingers," *J. Gen. Psychol.* 1, 353-359.
- Geldard, F. A. (1957). Adventures in tactile literacy. *American Psychologist*, Vol 12(3), 115-124.
- Gescheider, G. A., Bolanowski, S. J., and Verrillo, R. T. (1989). "Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency," *J. Acoust. Soc. Am.* 85, 2059-2064.
- Gescheider, G. A., Bolanowski, S. J., Verrillo, R. T., Arpajian, D. J., and Ryan, T. F. (1990). "Vibrotactile intensity discrimination measured by three methods," *J. Acoust. Soc. Am.* 87, 330-338.
- Gescheider, G. A., Valetutti, A. A., Padula, M. C., and Verrillo, R. T. (1992). "Vibrotactile forward masking as a function of age," *J. Acoust. Soc. Am.* 91, 1690-1696.
- Gick B, Derrick D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502-504.
- Goff, G. D. (1967). "Differential discrimination of frequency of cutaneous mechanical vibration," *J. Exp. Psychol.* 74, 294-299.
- Grant KW, Braida LD, Renn RJ. (1991) Single band amplitude envelope cues as an aid to speechreading. *The Quarterly Journal of Experimental Psychology, Section A*, 43(3):621-645.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. (Wiley, New York).
- Greenspan JD, Bolanowski SJ. (1996). The psychophysics of tactile perception and its peripheral physiological basis. In L. Kruger, ed. 1996. *Pain and Touch*. San Diego, CA: Academic. 2nd ed. pp. 25-103
- Heider F, Heider G. (1940) An experimental investigation of lipreading. *Psychol. Monogr.* 52, 124-153.
- Heming, J. E., and Brown, L. N. (2005). "Sensory temporal processing in adults with early hearing loss," *Brain Cognition* 59, 173-182.

- Hirsh, I. J., and Sherrick, C. E. (1961) "Perceived order in different sensory modalities," *J. Exp. Psychol.* 62, 423-432.
- Houde JF, Jordan MI. (1998). Sensorimotor adaptation in speech production. *Science.* 279, 1213–1216.
- House AS. (1961) "On Vowel Duration in English," *J. Acoust. Soc. Am.* Volume 33, Issue 9, pp. 1174-1178
- House AS, Fairbanks G. (1953) "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," *J. Acoust. Soc. Am.* Volume 25, Issue 1, pp. 105-113.
- Hudgins CV. (1934). A comparative study of the speech coordinations of deaf and normal subjects. *Journal of Genetic Psychology*, 44, 1–48.
- Ifukube T, Yoshimoto C. (1974). A sono-tactile deaf-aid made of piezoelectric vibrator array. *Journal of the acoustical society of Japan.* 30, 461-462.
- Ito T, Tiede M, Ostry DJ. (2009). Somatosensory function in speech perception. *Proc Natl Acad Sci.* 106(4): 1245-1248.
- Jeffers J., Barley M. (1977). *Speechreading (Lipreading)*. Charles C. Thomas Publisher, Springfield, IL.
- Knudsen, V.O. (1928). Hearing with the sense of touch. *J. Gen. Psychol.* 1:320-352.
- Kunisaki O, Fujisaki H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatics (University of Tokyo)*. 11:85-91.
- Ladefoged P., Maddieson I. (1996). *The sounds of the world's languages*. Wiley-Blackwell Ltd., Malden, MA.
- Lamore, P. J. J., Muijser, H., and Keemink, C. J. (1986). "Envelope detection of amplitude modulated high-frequency sinusoidal signals by skin mechanoreceptors," *J. Acoust. Soc. Am.* 79, 1082-1085.
- Lane H, Perkell JS. (2005) Control of Voice-Onset Time in the Absence of Hearing: A Review. *Journal of Speech, Language, and Hearing Research* Vol.48 1334-1343.
- Lazzaro J., Mead C. (1989). Circuit models of sensory transduction in the cochlea. In Mead, C. and Ismail, M. (eds), *Analog VLSI Implementations of Neural Networks*. Norwell, MA: Kluwer Academic Publishers, pp. 85-101.
- Levanen, S., and Hamdorf, D. (2001) "Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans," *Neurosci.Letters* 301, 75-77.
- Levitt, H. (1971). "Transformed up-down procedures in psychoacoustics," *J. Acoust. Soc. Am.* 49, 467-477.
- Liberman AM. (1996). *Speech: a special code*. Cambridge, MA: MIT Press.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. (1967). Perception of the speech code. *Psychological Review.* 74, 431-461.

- Lieberman AM, Delattre P, Cooper FS. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*. 65(4): 497-516.
- Lieberman AM, Mattingly IG. (1985). The motor theory of speech perception revised. *Cognition*. 21, 1-36.
- Luce PA, Charles-Luce J. (1985) "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," *J. Acoust. Soc. Am.* Volume 78, Issue 6, pp. 1949-1957
- Mack M. (1982) . "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production," *J. Acoust. Soc. Am.* Volume 71, Issue 1, pp. 173-178.
- MacLeod A, Summerfield Q. (1987) Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol.* 21, 131-141.
- Macmillan, N. A., and Creelman, C. D. (1990) *Detection Theory: A User's Guide*. (Cambridge University Press).
- Markram H, Gerstner W and Sjöström PJ (2011). A history of spike-timing-dependent plasticity. *Front. Syn. Neurosci.* 3:4.
- Marks, L. E., Girvin, J. P., O'Keefe, M. D., Ning, P., Quest, D. O., Antunes, J. L., and Dobelle, W. H. (1982). "Electrocutaneous stimulations III: The perception of temporal order," *Percept. Psychophys.* 32, 537-541.
- Moallem TM, Reed CM, Braida LD. (2010) "Measures of tactual detection and temporal order resolution in congenitally deaf and normal-hearing adults," *J Acoust Soc Am.* 127(6):3696-709.
- Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., and Wright, B. A. (2006) "Perceptual-learning evidence for separate processing of asynchrony and order tasks," *J. Neurosci.* 26, 12708-12716.
- Marslen-Wilson W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522-523.
- McClelland JL, Elman JL. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1): 1-86.
- McGurk H, MacDonald J. (1976). Hearing lips and seeing voices. *Nature*. 264: 746-748.
- Newman RS. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *J Acoust Soc Am.* 113(5): 2850-2860.
- Nicholas, J. G., and Geers, A. E. (2007). "Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss," *J. Speech Lang. Hear. Res.* 50, 1048-1062.
- Ohngren G, Rönneberg J, Lyxell B. (1990). Tactiling: Usable support system for speechreading? *STL-QPSR*, 31:69-77.
- Ohngren G. (1992). Touching voices with the Tactilator. *STL-QPSR*, 33:23-39.

Ouellet, C., and Cohen, H. (1999). "Speech and language development following cochlear implantation," *J. Neurolinguist.* 12, 271-288.

Pastore, R. E. (1983). "Temporal order judgment of auditory stimulus offset," *Percept. Psychophys.* 33, 54-62.

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, Zandipour M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *J Acoust Soc Am.* 116(4 Pt 1):2338-44.

Pickett JM, Pickett BH. (1963). Communication of speech sounds by a tactual vocoder. *J Speech Hear Res.* 10:207-22.

Plant G., Spens K-E. (1986). An experienced user of tactile information as a supplement to lipreading: An evaluative study. *STL-QPSR*, 1, 87-110.

Plant G, Gnosspeilius J, Levitt H. (2000). The Use of Tactile Supplements in Lipreading Swedish and English: A Single-Subject Study. *Journal of Speech, Language & Hearing Research.* 43(1): 172-183.

Porter RJ, Lubker JF. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage in speech. *Journal of Speech and Hearing Research*, 23, 593-602.

Rabinowitz, W. M., Houtsma, A. J. M., Durlach, N. I., and Delhorne, L. A. (1987). "Multidimensional tactile displays-identification of vibratory intensity, frequency, and contactor area," *J. Acoust. Soc. Am.* 82, 1243-1252.

Raphael LJ. (1972) "Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English," *J. Acoust. Soc. Am.* Volume 51, Issue 4B, pp. 1296-1303.

Rauschecker, J. P., and Shannon, R. V. (2002). "Sending sound to the brain," *Science*, 295, 1025.

Reed CM. (1995) Tadoma: An Overview of Research, in *Profound Deafness and Speech Communication*, K-E Spens and G Plant, Eds. Singular Publishing Group; New Ed edition (September 1995).

Reed, C. M., Durlach, N. I., and Braida, L. D. (1982a). "Research on tactile communication of speech: A review," *ASHA Monogr.* No. 20.

Reed, C. M., Durlach, N. I., and Braida, L. D. (1982b). "Analytic study of the Tadoma method: Identification of consonants and vowels by an experienced Tadoma user," *J. Speech Hearing Res.* 25, 108-116.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., Schultz, M. C. (1985). "Research on the Tadoma method of speech communication.," *J. Acoust. Soc. Am.* 77, 247-257.

Reed, C. M., Durlach, N. I., Delhorne, L. A., Rabinowitz, W. M., and Grant, K. W. (1989) "Research on tactual communication of speech: Ideas, issues, and findings," *Volta Rev. (Monogr.)* 91, 65-78.

- Reed CM, Delhorne LA, Durlach NI, Fischer SD. (1990) "A study of the tactual and visual reception of fingerspelling," *J Speech Hear Res.* Dec;33(4):786-97.
- Reed CM, Delhorne LA, Durlach NI, Fischer SD. (1995) "A study of the tactual reception of sign language," *J Speech Hear Res.* Apr;38(2):477-89.
- RLE Quarterly Progress Report, No.13 through No. 18 (1949-50). Communications Research: Sensory replacement project (FELIX). Research Laboratory of Electronics, MIT, Cambridge, MA.
- Rosen SM, Fourcin AJ, Moore BCJ. (1981). Voice pitch as an aid to lipreading. *Nature* 291, 150-152.
- Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L., Bolanowski, S. J., Jr. (1977). "Vibrotactile frequency for encoding a speech parameter," *J. Acoust. Soc. Am.*, 62, 1003-1012.
- Schiff, W., and Dytell, R. S. (1972). "Deaf and hearing children's performance on a tactual perception battery," *Percept. Motor Skill* 35, 683-706.
- Secker-Walker HE, Searle CL. (1990) Time-domain analysis of auditory-nerve-fiber firing rates. *J Acoust Soc Am* 1990 Sep; 88(3): 1427-36.
- Sherrick, C. E. (1964). "Effects of double simultaneous stimulation of the skin," *Am. J. Psychol.* 77, 42-53.
- Sherrick, C. E. (1970). "Temporal ordering of events in haptic space," *IEEE Trans Man- Machine Syst.* MMS-11, 25-28.
- Shore, D. I., Spry, E., and Spence, C. (2002). "Confusing the mind by crossing the hands," *Cognitive Brain Res.* 14, 153-163.
- Sparks DW, Kuhl PK, Edmonds AE, Gray GP. (1978). Investigating the MESA (multipoint electro tactile speech aid): the transmission of segmental features of speech. *J Acoust Soc Am.* 63(1):246-57.
- Stenquist, G. (1974). *The Story of Leonard Dowdy: Deaf-Blindness Acquired in Infancy.* Watertown, Massachusetts: Perkins School for the Blind.
- Stevens KN. (1999). *Acoustic Phonetics.* The MIT Press, Cambridge, MA.
- Stevens KN. (2002). Toward a model of lexical access based on acoustic landmarks and distinctive features. *J Acoust Soc Am.* 111(4): 1872-1891.
- Stremmer FG. (1990). *Introduction to Communication Systems, 3rd Edition.* Addison-Wesley Publishing Co., New York.
- Tamir TJ (1989) *Characterization of the Speech of Tadoma Users.* B.S. Thesis. Massachusetts Institute of Technology, Cambridge, MA.
- Tan, H. Z. (1996). "Information transmission with a multi-finger tactual display," Ph.D dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Tan, H. Z., and Rabinowitz, W. M. (1996). "A new multi-finger tactual display, " in *Proceedings of the Dynamic Systems and Control Division, DSC-Vol. 58,* pp. 515-522.

- Taylor, B. (1978). "Dimensional redundancy in the processing of vibrotactile temporal order," Ph.D. dissertation, Princeton University, Princeton, NJ.
- Taylor M. M., Forbes S. M., and Creelman C. D. (1983). "PEST reduces bias in forced choice psychophysics," *J. Acoust. Soc. Am.* 74, 1367-1374.
- Traunmüller H. (1994). Conventional, biological and environmental factors in speech communication: A modulation theory. *Phonetica*. Vol. 51: 170–183. [see also author's online review: <http://www.ling.su.se/staff/hartmut/module.htm>]
- Tyler, R. S., Gantz B. J., Woodworth G. G., Pakinson A. J., Lowder M. W., Schum L. K. (1996). "Initial independent results with the Clarion cochlear implant," *Ear Hear* 17, 528-536.
- Uchanski, R. M., and Geers, A. E. (2003). "Acoustic characteristics of the speech of young cochlear implant users: A comparison with normal-hearing age-mates," *Ear Hear*. 24, 90S-105S.
- Valente M., Fernandez E., Monroe H. (2010). *Audiology Answers for Otolaryngologists: A High-Yield Pocket Guide*. New York: Thieme Medical Publishers, Inc.
- Van Doren, C. L., Gescheider, G. A., and Verrillo, R. T. (1990). "Vibrotactile temporal gap detection as a function of age," *J. Acoust. Soc. Am.* 87, 2201-2206.
- Vernon, M. (1969) "Usher's syndrome -- deafness and progressive blindness. Clinical cases, prevention, theory and literature survey," *Journal of Chronic Disorders* 22(3), 133-151.
- Verrillo, R. T. (1963). "Effect of contactor area on the vibrotactile threshold," *J. Acoust. Soc. Am.*, 35, 1962-1966.
- Verrillo, R. T. (1965). "Temporal summation in vibrotactile sensitivity," *J. Acoust. Soc. Am.* 37, 843-846.
- Verrillo, R. T. (1966). "Vibrotactile thresholds for hairy skin," *J. Exp. Psychol.*, 72, 47-50.
- Verrillo, R. T. (1971). "Vibrotactile thresholds measured at the finger," *Perception and Psychophysics*, 9, 329-330.
- Verrillo, R. T., Gescheider, G. A., Calman, B. G., and Van Doren, C. L. (1983). "Vibrotactile masking: Effects of one and two-site stimulation," *Percept. Psychophys.* 33, 379-387.
- Verrillo, R. T., and Gescheider, G. A. (1992). "Perception via the sense of touch," In *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr Publishers, London), pp. 1–36.
- Vivian, R. M. (1966). *The Tadoma Method: A Tactual Approach to Speech and Speechreading*. The *Volta Review*, December, 733-737.
- Walden BE, Prosek RA, Montgomery AA, Scherr CK, Jones CJ. (1977) Effects of training on the visual recognition of consonants. *J. Speech Hear. Res.* 20, 130-145.
- Walden BE, Erdman SA, Montgomery AA, Schwartz DM, Prosek RA. (1981). Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech and Hearing Research*, 24:207-216.

Wilson, B. S., and Dorman M. F. (2008). "Cochlear implants: Current designs and future possibilities," *J. Rehabil. Res. Dev.* 45, 695-730.

World Health Organization. (2006). "Deafness and hearing impairment," Fact Sheet No. 300, April 2010. Available at <http://www.who.int/mediacentre/factsheets/fs300/en/> [last viewed: 7/28/11].

Yale CA. (1892; 1946 reprint). *Formation and development of elementary English sounds*. The Clarke School for the Deaf, Northampton, MA.

Yuan, H. F. (2003). "Tactual display of consonant voicing to supplement lipreading". Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (2004a). "Envelope-onset asynchrony as a cue to voicing in initial English consonants," *J. Acoust. Soc. Am.* 116, 3156–3167.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (2004b). "Envelope offset asynchrony as a cue to voicing in final English consonants," *J. Acoust. Soc. Am.* 116, 2628 (A). [Poster presentation]

Yuan, H. F., and Reed, C. M. (2005). "A tactual display of envelope-offset asynchrony as a cue to voicing of consonants in the final position," Talk presented at Eighth International Sensory Aids Conference, Portland, ME, May 12, 2005.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (2005a). "Temporal onset-order discrimination through the tactual sense," *J. Acoust. Soc. Am.* May 117, 3139-3148.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (2005b) "Tactual display of consonant voicing as a supplement to lipreading," *J. Acoust. Soc. Am.* 118, 1003-1015.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (2006) "Temporal onset-order discrimination through the tactual sense: Effects of frequency and site of stimulation," *J. Acoust. Soc. Am.* 120, 375-385.