Veterans Administration

Journal of Rehabilitation Research and Development Vol. 25 No. 4
Pages 45–52

Preliminary results of speech-reception tests obtained with the synthetic Tadoma system

D.F. LEOTTA, W.M. RABINOWITZ, C.M. REED, and N.I. DURLACH
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract—In the Tadoma method of speech reception used by some deaf-blind individuals, speech is understood by placing a hand on the face of the talker and feeling certain mechanical actions of the face associated with speech production. The synthetic Tadoma system is a computer-driven artificial face that simulates these mechanical actions. This paper reports some preliminary data on the discrimination of nonsense syllables with the synthetic system. Although further work is required to produce an accurate simulation, the results suggest that such a goal is indeed achievable.

Key words: *deaf-blind, tactual speech reception.*

INTRODUCTION AND BACKGROUND

Tadoma is a method of tactile speech communication that has evolved within the deaf-blind community. In this method, speech is received by placing a hand on the face of the talker and monitoring facial actions associated with the speech production process (**Figure 1**). Evaluations of Tadoma for both experienced deaf-blind Tadoma readers and naive normally-sensed subjects have been described by Norton *et al.* (5) and Reed *et al.* (6–12). According to these results, experienced readers can achieve almost normal communication with this method (e.g., 80 percent keywords correct on "everyday" sentences spoken at a rate of 3 syllables per second). Possible factors underlying the remarkable performance achieved with this method by experienced users include: a) the nature of the Tadoma display—a talking face; b) the large amount of training these users have had; and, c) the use of the hand for receiving the tactile input. Two features of the display that are thought to be important are its perceptual multidimensionality (4) and its direct relation to the articulation process of speech production (3). The idea that deaf-blind Tadoma readers have exceptional tactile sensitivity is contradicted by results showing that their ability to discriminate or identify isolated nonsense syllables using Tadoma is no better than that of naive normally-sensed subjects with only modest amounts of training.

In order to explore various research questions that have arisen in the study of Tadoma, we are developing a synthetic Tadoma system. The initial objective of this work is to create a system that provides an accurate simulation of Tadoma, i.e., provides speech perception performance that is the same as that obtained with natural Tadoma. Once this objective has been achieved, the system will then be used to study how performance is affected by various transformations that cannot be realized with natural Tadoma. By examining how performance depends on selected characteristics of the input information and/or the display, we will be able to identify those features of Tadoma responsible for its unusual success. This knowledge will then provide important background for the development of im-

Address all correspondence and reprint requests to: N.I. Durlach, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Ave., Room 36-709, Cambridge, MA 02139.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Figure 1.
The Tadoma method in use. Two deaf-blind, experienced users of the method (LD, left, and RB, right) are communicating with each other and with a hearing and sighted member of our research team (ND, center). [Photo courtesy of H. Durlach; appeared previously in (12)]

proved tactile aids in which the tactile stimulation patterns are driven by the acoustic waveforms of speech rather than by physical contact with the talker.

The synthetic Tadoma system is composed of an artificial mechanical face that is driven by computer-controlled facial-action signals which are derived from sensors mounted on a real talking face (**Figure 2**). The facial actions included in this system consist of those actions that seem to provide the primary cues for understanding speech via Tadoma (movements of the lips and jaw, vibration of the larynx, and airflow from the mouth). The specific facial actions represented in the synthetic Tadoma system, the sensors used to measure them from a talker, and their representation in the artificial face, are summarized in **Table 1**. Further details on Tadoma and on the synthetic Tadoma system are available in Reed *et al.* (12) and Leotta (2). In the present paper, we report some preliminary results on the perception of speech segments via the synthetic Tadoma system.

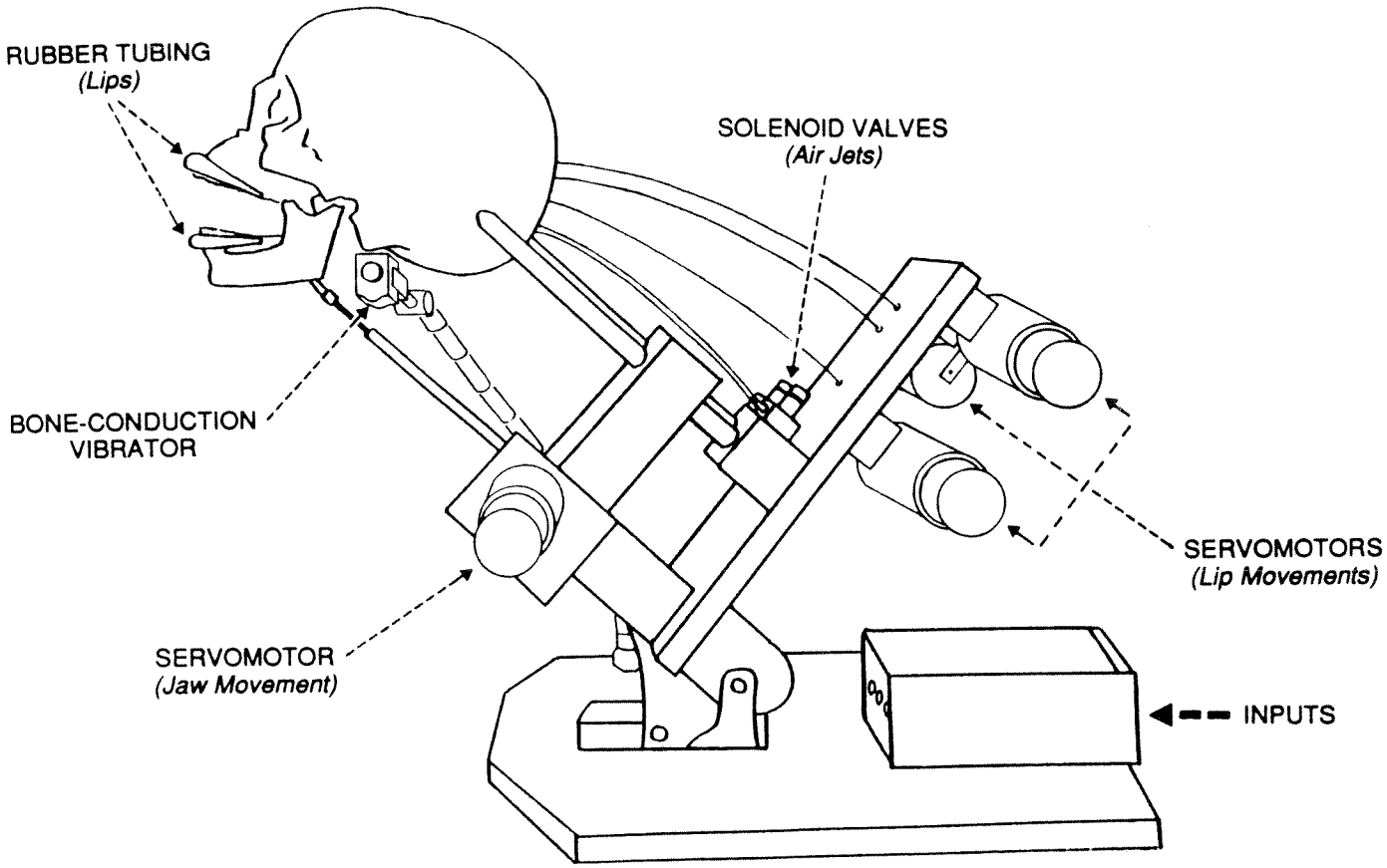

Figure 2.
Schematic illustration of artificial-face display.

PRELIMINARY DISCRIMINATION EXPERIMENTS

The experimental procedure was similar to the one used previously by Reed *et al.* (1978) to study reception of natural Tadoma by naive, normally-sensed listeners (7). (In the following discussion, that study will be referred to as the Reed *et al.* 1978 study, and the study which is the subject of this paper will be referred to as the current study.) The speech materials consisted of nonsense syllables [vowels (V) and consonants (C) in CVC and CV contexts] and the experimental testing paradigm was roving-contrast ABX discrimination with trial-by-trial feedback. Among the differences in the 2 studies was the use of recorded materials and the longer trial times in the current study. In a given test list, the speaker and context were held fixed, but the pair of items to be discriminated, that is, the contrast, varied randomly from trial to trial (hence the word "roving"). Furthermore, 2 tokens (that is, 2 utterances of the same speech element by the same talker) were available for each item, and the same token was never used twice within a trial. Thus, for example, on a trial of the form ABA, the token for the second A was distinct from the token for the first A. All contrasts were minimal in the sense that A and B differed in only one phonetic feature.

Two subjects, DL and CR, were tested by two speakers, DB and WR. A complete set of materials (including the 2 utterances per item) were obtained once for speaker DB and twice for speaker WR (thus referred to as speaker WR1 and speaker WR2). Subject DL had no previous experience with either form of Tadoma (natural or synthetic), while subject CR had experience only with natural Tadoma. On all tests, blindfolds and masking noise were used to eliminate all non-tactual cues (e.g., sounds produced by the actions of the motors, air valves, and vibrator).

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 1.
Parameters of the synthetic Tadoma system

Facial Action	Sensor	Artificial Face Representation
Laryngeal vibration	Throat microphone	Vibrator
Oral airflow	Anemometer in front of the mouth	Four-level airflow via high-speed valves
Articulatory movements: (a) Jaw height (b) Lower lip height (c) Lower lip protrusion (d) Upper lip protrusion	Cantilevered strain gauges	Position controlled actuators

Vowels

Experiments on vowel discrimination tested 11 vowel pairs in 3 CVC contexts (/h/-V-/d/, /g/-V-/d/, and /b/-V-/d/). The pairs contrasted the features round, tense, back, high, low (1). A given test list included 5 presentations of each of the 11 contrasts (all in the same context, all uttered by the same speaker) ordered randomly.

A comparison of the results of these tests in the current study, using synthetic Tadoma, and the results of the previous tests in the Reed *et al.* 1978 study, using natural Tadoma, are shown in **Figure 3**. The results on natural Tadoma are for the context /h/-V-/d/, the only context tested in the Reed *et al.* 1978 study, and were determined by averaging the results (Reed *et al.* 1978, column 2 of Table 2) obtained with the roving contrast paradigm over the 2 subjects tested (approximately 25 trials per contrast per subject), one of whom also participated in the current study (CR). The results on synthetic Tadoma give the scores for DL averaged over tests with each of the 3 speakers and 3 contexts (9 different experiments with 50 trials per contrast per experiment). The scores for DL in these 9 conditions (averaged over the 11 vowel contrasts) varied from 63 percent to 80 percent with an overall average of 71 percent.

The results for CR with synthetic Tadoma are not included in **Figure 3**, because the amount of data obtained with CR was so small (only speaker DB and only the /h/-V-/d/ context were tested with CR). It should be noted, however, that: a) CR's score on synthetic Tadoma was less than DL's—66 percent compared to 71 percent; b)DL's score on the /h/-V-/d/ context was slightly smaller than on the other contexts—67 percent versus 72 percent and 74 percent; and, c) CR's score on natural Tadoma in the Reed *et al.* 1978 study was better than the score of the other subject tested in that study—72 percent versus 68 percent. Thus, if the comparison included CR's results on synthetic Tadoma, or the comparison was restricted to just CR's results, or just the /h/-V-/d/ results, the relative performance with the synthetic system would be slightly worse than it appears in **Figure 3**.

Despite these qualifications, and despite the possible biasing of the comparison by other factors that are difficult to quantify (e.g., the live-voice presentations versus the recorded 2-token presentations), it appears that the overall performance of the 2 systems on the vowel tests is equivalent, at least to a first order approximation. Furthermore, as shown in **Figure 3**, where the contrasts are ordered according to the discrimination difficulty in synthetic Tadoma, there is a rough similarity between the methods in the relative degree of discrimination difficulty among the contrasts. As one might expect, for both methods, the pairs with poor scores tend to contain vowels that differ primarily in tongue position rather than lip position. Note, however, that there seems to be less variation among the scores for the synthetic case than for the natural case and, in particular, that only for the natural case do any of the scores exceed 90 percent.

Consonants

Experiments on consonant discrimination tested 38 contrasts in each of 3 CV contexts (C-/a/, C-/i/, and C-/u/). A given test list included 1 presentation of each contrast (all in the same context, all uttered by the same speaker) ordered randomly.

A comparison of the results of these tests in the current study, using synthetic Tadoma, and the results of the previous tests in the Reed *et al.* 1978 study, using natural Tadoma, are shown in **Table 2**. The results on natural Tadoma in columns 1–3 show the individual scores for the subjects CR and SR obtained from the Reed *et al.* 1978 study (CV column in Table 3), as well as the average over these 2 subjects. For each subject, the scores are based on roughly 30 trials per contrast from an experiment in

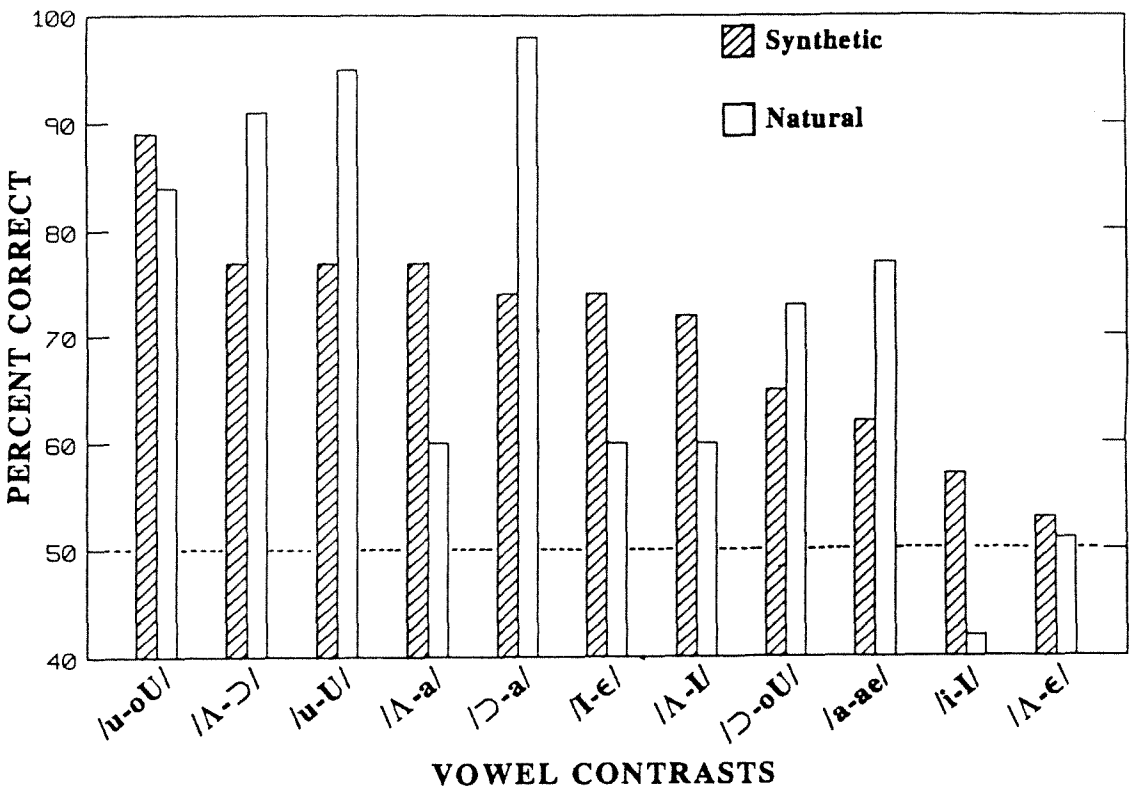

Figure 3.
Vowel discrimination scores for the synthetic system and natural Tadoma. Synthetic system: Present study, subject DL, 450 trials/contrast, results averaged over the 3 speakers (WR1, WR2, and DB) and 3 contexts (/h/-V-/d/, /g/-V-/d/, and /b/-V-/d/). Natural Tadoma: Roving stimulus scores from Reed *et al.* (1978, **Table 2**), results averaged for subjects SR and CR, approximately 25 trials/contrast/subject in the context /h/-V-/d/. Contrasts are arranged in decreasing order of performance on the synthetic system.

which only one talker was used, but the vowel context varied randomly from trial to trial over the set C-/a/, C-/i/, and C-/u/. Subsequent processing showed approximately the same scores for the 3 contexts. (Although the tests were run with CV's and VC's intermixed randomly within each list, the scores and number of trials cited refer only to the CV presentations.) The results on synthetic Tadoma in columns 5–7 show the individual scores for subjects CR and DL for talker WR1 and context C-/a/ based on 30 trials per contrast, as well as the average of these scores over these 2 subjects. The results in column 8 show the scores for subject DL averaged over all 3 contexts for talker WR1 and over all 3 talkers for context C-/a/. The score obtained by DL for each of these 5 combinations of talker and context, averaged over all 38 contrasts, was in the range 63 percent to 68 percent, except for the combination (WR1, C-/u/) when the score was 57 percent. The number of trials for each combination was 30 trials per contrast, leading to a total of 5700 trials for subject DL. Note, finally, that in each column of **Table 2** the average results for each major class (voicing, place, and manner) were obtained by weighting the scores according to the number of contrasts tested.

Unlike the results on vowels, the results on consonants seem distinctly worse for the synthetic case. Although the average results for manner distinctions are nearly the same, the average results for the voicing distinctions are much worse in the synthetic case than in the natural case (with the differential results for the place distinctions falling in-between).

Voicing distinctions are essentially absent with the synthetic system: No score shown in columns 5–8 of **Table 2** exceeds 58 percent correct. Within the place contrasts, only the semivowels, liquids, and glides lead to comparable scores with the 2 systems: Plosives, fricatives, and especially nasals, are much

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 2.
Consonant discrimination scores

		NATURAL				SYNTHETIC				
	Speaker	SR	CR	-	WR	WR1	WR1	WR1	ALL	WR1
	Context	/a,i,u/	/a,i,u/	/a,i,u/	/a/	/a/	/a/	/a/	ALL	/a/
	Trials/Pair	30	30	60	10	30	30	60	150	10
	Column	1	2	3	4	5	6	7	8	9
	Subject	CR	SR	AVG	DL	CR	DL	AVG	DL	DL-M
A.	**Voicing contrasts for cognate pairs**									
	Plosives /p-b/, /t-d/, /k-g/	71	68	70	73	58	44	51	55	73
	Fricatives /f-v/, /s-z/, /sh-zh/	72	72	72	90	53	58	56	58	87
	Affricates /ch-j/	61	65	63	60	53	57	55	55	80
	Average	**70**	**69**	**70**	**78**	**55**	**52**	**54**	**56**	**82**
B.	**Place contrasts within phoneme classes**									
	Plosives /p-t/, /p-k/, /t-k/	97	87	92	87	70	71	71	64	60
	Fricatives /f-s/, /f-sh/, /s-sh/	88	82	85	77	78	66	72	58	70
	Nasals /m-n/	100	100	100	100	70	43	57	51	60
	Semivowels, Liquids, Glides /r-w/, /r-y/, /r-hw/, /w-hw/, /l-y/, /l-h/, /y-h/	72	75	74	69	75	76	76	64	76
	Average	**83**	**81**	**82**	**77**	**74**	**70**	**73**	**62**	**70**

worse with the synthetic system. Finally, note that although the overall average for manner contrasts is nearly the same in the 2 cases, there is only limited consistency for individual classes (e.g., plosive-nasal discrimination is much more difficult with synthetic Tadoma than with natural Tadoma).

INITIAL MODIFICATION OF THE DISPLAY

The most obvious and important deficiency of the synthetic Tadoma system indicated by the above tests is the loss of voicing information. After determining from supplementary tests that subject DL, like CR, can discriminate voicing with natural Tadoma, an attempt was made to modify the synthetic system to better display voicing. (DL's results on natural Tadoma with speaker WR and the context C-/a/, based on 10 trials for each of 38 contrasts, are shown in column 4 of **Table 2**).

In the original display, the voicing signal was displayed to only 1 finger. The display was modified so that the voicing signal was displayed to all 4 fingers (increasing the similarity with natural

Table 2. *(Continued)*
Consonant discrimination scores

		NATURAL				SYNTHETIC				
	Speaker	SR	CR	–	WR	WR1	WR1	WR1	ALL	WR1
	Context	/a,i,u/	/a,i,u/	/a,i,u/	/a/	/a/	/a/	/a/	ALL	/a/
	Trials/Pair	30	30	60	10	30	30	60	150	10
	Column	1	2	3	4	5	6	7	8	9
	Subject	CR	SR	AVG	DL	CR	DL	AVG	DL	DL-M
C.	**Manner contrasts for similar places**									
	Plosive-Fricative /p-f/, /t-s/, /k-sh/	79	87	83	80	70	78	74	69	70
	Plosive-Nasal /p-m/, /b-m/, /d-n/	85	78	82	60	48	64	56	62	57
	Plosive-Affricate /t-ch/, /k-ch/	76	84	80	85	92	92	92	80	90
	Plosive-Glide /p-w/, /b-w/	84	83	84	75	73	73	73	69	85
	Plosive-Sonorant /t-l/	91	68	80	100	93	57	75	61	70
	Fricative-Affricate /s-ch/, /sh-ch/	55	64	60	70	77	67	72	68	70
	Nasal-Glide /m-w/	82	75	78	70	60	77	69	64	90
	Nasal-Sonorant /n-l/	81	63	62	70	87	53	69	64	70
	Fricative-Fric. Glide /zh-h/	95	74	84	100	97	93	95	79	100
	Plosive-Fric. Glide /g-h/	81	70	74	90	83	83	83	64	90
	Average	**78**	**77**	**78**	**77**	**74**	**74**	**74**	**68**	**76**

Tadoma) by attaching to the vibrator a metal bar that extended from the throat region to the zygomatic arch. Also, minor modifications were made to the airflow system, to the movement files, and to certain durational aspects of the experimental paradigm. However, these other changes are not believed to have influenced performance nearly as much as the addition of the metal bar.

Following the modifications, the tests were repeated on DL, again using speaker WR1, context C-/a/, and 10 trials for each of 38 contrasts. The results of these tests are shown in column 9 of **Table 2**. Not only do these scores indicate dramatically improved performance on voicing contrasts (from 52 percent to 82 percent), but no loss of performance is evident on place contrasts (which remained at 70 percent) or manner contrasts (which increased from 74 percent to 76 percent). Compared to natural Tadoma, performance with the synthetic system now seems to suffer only from a modest decrement in place information, as shown, for example, by comparing the averages in column 9 to those in column 4.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

CONCLUDING REMARKS

Our preliminary results suggest that the synthetic Tadoma system can provide at least a rough simulation of natural Tadoma. They also indicate that rather minor modifications of the system, such as inclusion of the vibratory bar, can have a substantial effect on performance.

In future work, we intend to explore other modifications of the display in an attempt to further improve the quality of the simulation. Detailed comparisons of the errors made with the 2 methods, as well as the subjective impressions of individuals tested with the 2 methods, suggest that the synthetic system needs additional modifications to obtain more complete lip closure and to represent airflow information more adequately.

Assuming that these modifications lead to a simulation that is adequate according to the above types of discrimination tests, we will continue our evaluation with studies of segment identification (using both naive subjects and experienced Tadoma users) and sentence reception (using only experienced Tadoma users).

If the synthetic system passes these more comprehensive simulation tests, we will then exploit synthetic Tadoma as a research tool by selectively modifying the input signals and/or display (in ways that cannot be done with natural Tadoma) and observing how performance is influenced by these modifications. In this way, we will: 1) test various hypotheses about the roles played by different physical cues in the system; 2) explore the differential effects of modifying the input information while maintaining a constant display versus maintaining constant input information while modifying the display; and 3) examine the sensitivity of performance to signal bandwidths and rates.

Finally, we will attempt to modify the synthetic Tadoma system so that its performance exceeds that of natural Tadoma (e.g., by displaying some representation of tongue position), thus creating a "Super Tadoma" display and establishing a new standard of performance for the tactile communication of speech.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (Grant BNS 84-17817).

REFERENCES

1. **Chomsky N, Halle M:** *The Sound Pattern of English,* New York: Harper and Row, 1968.
2. **Leotta DF:** Synthetic tadoma: Tactile speech reception from a computer-controlled artificial face. MS thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1985.
3. **Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M:** Perception of the speech code. *Psychol Rev* 74:431-461, 1967.
4. **Miller GA:** The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63:81-97, 1956.
5. **Norton SJ, Schultz MC, Reed CM, Braida LD, Durlach NI, Rabinowitz WM, Chomsky C:** Analytic study of the Tadoma method: Background and preliminary results. *J Speech Hear Res* 20:574-595, 1977.
6. **Reed CM, Durlach NI, Braida LD, Norton SJ, Schultz MC:** Experimental results on Tadoma. In *Papers from the Research Conference on Speech-Processing Aids for the Deaf,* J.M. Pickett (Ed.), 145-147. Washington, DC: Gallaudet College, 1977.
7. **Reed CM, Rubin SI, Braida LD, Durlach NI:** Analytic study of the Tadoma method: Discrimination ability of untrained observers. *J Speech Hear Res* 21:625-637, 1978.
8. **Reed CM, Doherty MJ, Braida LD, Durlach NI:** Analytic study of the Tadoma method: Further experiments with inexperienced observers. *J Speech Hear Res* 25:216-223, 1982a.
9. **Reed CM, Durlach NI, Braida LD:** Research on tactile communication of speech: A review. *Am Speech-Language-Hearing Assoc (ASHA) Monograph No. 20* 1-23, 1982.
10. **Reed CM, Durlach NI, Braida LD, Schultz MC:** Analytic study of the Tadoma method: Identification of consonants and vowels by an experienced Tadoma user. *J Speech Hear Res* 25:108-116, 1982.
11. **Reed CM, Rabinowitz WM, Durlach NI, Braida LD:** Tactile perception of speech. *Research Laboratory of Electronics Progress Report 126,* MIT, Cambridge, MA 1984.
12. **Reed CM, Rabinowitz WM, Durlach NI, Braida LD, Conway-Fithian S, Schultz MC:** Research on the Tadoma method of speech communication. *J Acoust Soc Am* 77:247-257, 1985.