

Bimodal Speech Perception in Infant Hearing Aid and Cochlear Implant Users

Brittan A. Barker, MA; J. Bruce Tomblin, PhD

Objectives: To determine the feasibility of replicating prior bimodal perception findings with hearing-impaired infants during their preimplant, hearing aid trial, and postimplant experiences; secondarily, to determine the point in development at which these infants were able to match phonetic information in the lips and voice for the vowels /a/ and /i/.

Methods: A total of 10 infants with hearing loss, aged 4 to 24 months, were assessed at least once prior to cochlear implantation and previous to implant stimulation. The Split-Screen Preferential Looking Procedure was used to evaluate the bimodal perception skills of these infants.

Results: Examples of individual bimodal perception data and preliminary group data are presented. A difference in performance across preimplant and postimplant test sessions was noted for the individuals and the group.

Conclusion: These data provide evidence that the infants' audibility levels were improved by their cochlear implants, which may have contributed to their evolving ability to match phonetic information in the lips and voice.

Arch Otolaryngol Head Neck Surg. 2004;130:582-586

THE AGE AT WHICH INFANTS undergo surgery for cochlear implants (CIs) has declined within recent years. With this decline, there has arisen a need for systematic research regarding the effects of implantation on infants and the development of their speech and language skills. Decades of research exploring speech perception have already shown that infants with normal hearing thresholds have the capabilities to store information about the acoustic signal and begin to learn from it even before they are born.¹⁻³ It is intended that the present study serve as an impetus to map the development of similar auditory skills in infant CI users and ultimately enable the field to use the present data as an indicator of speech perception skills in these young children.

Twenty-two years ago, Kuhl and Meltzoff⁴ began the investigation of infant speech perception as a multimodal process. Via a preferential looking procedure, the investigators examined the abilities of normal-hearing 4.5-month-old infants to match vowel information in a talker's lips and voice. The infants were presented with 2 synchronized films placed side by side. One film showed a woman articulating /i/, and the second film showed

the same woman articulating /a/. Simultaneously, the infants were presented with an audio stimulus composed of a woman saying /a/ or /i/. It was predicted that if the infants spent significantly more time looking in the direction of the film that corresponded to the audio stimulus (ie, the film of the woman articulating /a/ when the infant heard /a/), he or she was able to detect equivalent phonetic information in the talker's lips and voice.

The predictions of Kuhl and Meltzoff⁴ held true—the infants spent significantly more time looking at the target video that correctly matched the vowel sound presented via the loudspeaker, which indicates that despite a relative lack of experience producing and perceiving speech, these infants appeared to have bimodal perception already in place. Subsequent research has continued to support the idea that infants with normal hearing have an intermodal representation of speech.⁵⁻⁸ In fact, Patterson and Werker⁹ recently used digital technology to create dynamic audio and video stimuli and successfully replicated the results of Kuhl and Meltzoff with infants as young as 2 months.

A recent study of CI users, aged 4 to 8 years, suggests that pediatric CI users also appear to have an intermodal representa-

From the Departments of Otolaryngology–Head and Neck Surgery (Ms Barker) and Speech Pathology and Audiology (Dr Tomblin), University of Iowa, Iowa City. The authors have no relevant financial interest in this article.

tion of speech. Lachs and colleagues¹⁰ administered the Common Phrases Test¹¹ to evaluate speech perception skills under “auditory-alone and audio-visual” conditions. The results of the study demonstrated a significant enhancement in the children’s speech perception when the sentences were presented with both auditory and visual cues vs only auditory cues, thus suggesting that the pediatric CI users also perceive speech bimodally despite a lack of auditory input early in life.

Given the combination of bimodal perception data previously gathered from very young children with normal hearing⁹ and older children with CIs,¹⁰ it stood to reason that an evaluation of the bimodal perception of infant CI users was an appropriate place to begin a developmental investigation of these infants’ speech perception skills. The objectives of the present study were as follows: (1) to build on the work of Patterson and Werker^{9,12} and determine the feasibility of replicating their prior bimodal perception findings with infants who use hearing aids and/or CIs and (2) to determine the point in development at which infant CI users were able to match phonetic information in the lips and voice.

METHODS

DESIGN

This is an ongoing longitudinal study that approximates a multiple-baseline design. Normal-hearing infants were assessed at age 3 to 5 months, and hearing-impaired infants were assessed at least once prior to cochlear implantation and previous to CI stimulation (age 5-24 months).

PARTICIPANTS

A total of 10 children (3 girls) with profound, bilateral, sensorineural hearing loss (SNHL) participated in this study. Two infants failed to complete a minimum of 2 test sessions, so their data were excluded from the final analyses. In all cases, SNHL was identified by age 3 months, and all infants were born to hearing parents. The infants’ ages at the time of initial stimulation of their CI devices ranged from 11 months 5 days to 20 months 7 days, with an average (SD) age at initial stimulation of 15 months 11.7 days (2.62 months 6.70 days). All participants were observed longitudinally as part of a comprehensive CI center study.

Eight families described their primary mode of communication with their infants as total communication; 2 families described their primary mode of communication as oral/aural. American English was the primary language spoken in each infant’s home (ie, English was spoken more than 50% of the time in the infant’s listening environment). All infants’ cognitive abilities appeared to be within the normal limits as set out in the Bayley Scales of Infant Development II mental subscale.¹³ The infants had no known visual abnormalities.

SPLIT-SCREEN PREFERENTIAL LOOKING PROCEDURE

The Split-Screen Preferential Looking Procedure (SPLP) is a reliable paradigm (G. J. Hollich, PhD; K. Hirsh-Pasek, PhD; and R. M. Golinkoff, PhD, unpublished data, 2003) with a history in the field of cognitive development and often used to evaluate speech perception and language development in children aged 2 months¹⁴ to 36 months.¹⁵ The SPLP is specifically designed to determine if infants show a consistent preference for

a video event that is related to an acoustic stimulus. The index of preference is the difference in the length of the infant’s looking time at the 2 different kinds of visual stimuli over the test trial series. The resulting data are used to make inferences about various aspects of infants’ speech perception and language development. The SPLP was used to evaluate the bimodal perception skills of the infants in the present experiment.

Video Stimuli

All stimuli in this experiment were modeled after those items used in Patterson and Werker.¹² A woman was filmed articulating /a/ and then /i/; the final edited stimuli yielded 2 videos in which the model’s articulations were matched in duration while her head remained motionless and her eyes remained focused at midline.

Audio Stimuli

A different female native speaker of American English was chosen to record the audio stimuli. This ensured that no idiosyncratic aspects of the audio recording would facilitate the pairing of the audio and visual stimuli. The final audio stimuli set consisted of 3 articulations of /a/ and 3 articulations of /i/ spoken in infant-directed speech. The Cool Edit 2000 software package (Adobe Systems Inc, Scottsdale, Ariz, 2003) was used to edit each of the audio recording trios to make certain that there was no difference in amplitude levels between the stimuli sets.

Bimodal Stimuli

Cinestream 3.1 software (Media100 Inc, Marlboro, Mass, 2001) was used to ultimately combine the video tracks to form two 27-second-long loops of 9 articulations. The /a/ and /i/ video loops were then edited onto 1 screen to create a simultaneous side-by-side display of the video loops (**Figure 1**). The software was also used to combine and synchronize the simultaneous video loops with each of the audio tracks.

Apparatus

The SPLP setup was housed in a double-walled sound booth. A 52-inch television monitor was located in the front of the booth, and a video camera was mounted above the monitor. The orientation of each infant’s eyes was recorded via this video camera, and an experimenter was able to view the session via a video monitor outside the booth. The booth was lined with curtains; thus, only the monitor screen and the video camera’s lens were visible (**Figure 2**).

Procedure

The infant was seated on the caregiver’s lap in front of the monitor; the experimenter was seated outside of the booth. The caregiver wore a pair of glasses with opaque lenses. All trials began with a flashing red square in the middle of the monitor to capture the infant’s attention. Experimental sessions consisted of a familiarization phase and a test phase.

The familiarization phase consisted of silent trials in which the infant was introduced to the video images and their respective locations. The test phase consisted of the same video images presented in the familiarization phase; however, during the test phase, the infant was presented with speech stimuli played over the television’s central loudspeakers, and the 2 video images were presented simultaneously. The stimuli for each trial continued to play until its completion. Sound presented, left-right positioning of the 2 videos, and order of familiarization were counterbalanced.



Figure 1. Split-Screen Preferential Looking Procedure video still image of the stimuli for the present bimodal perception experiment. A, The talker articulating /i/; B, the same talker articulating /a/. The video stimuli were presented in synchrony in combination with either an /a/ or /i/ audio stimulus that was played over the television's central loudspeakers.

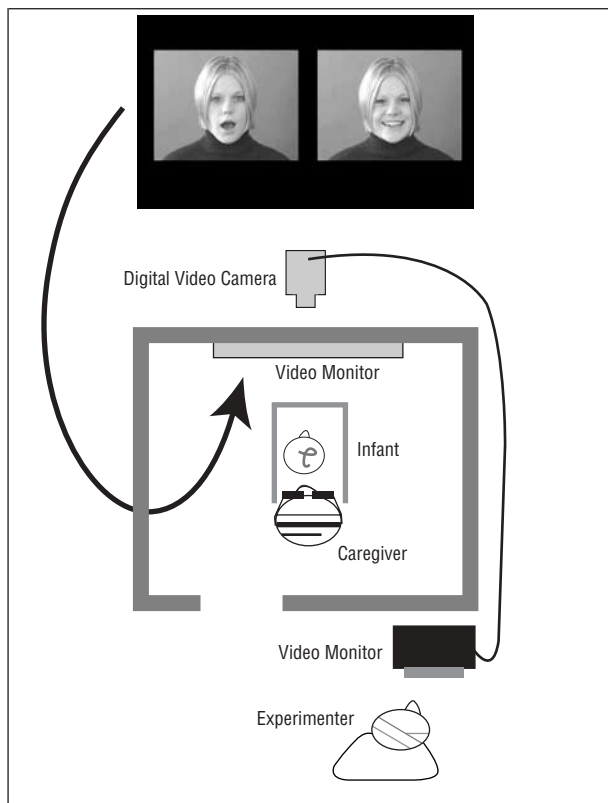


Figure 2. The digital video still image of the stimuli used in the experiment along with a schema of the Split-Screen Preferential Looking Procedure setup.

Coding

The infants' test sessions were coded off-line in accordance with the coding method for the SPLP used by Hollich.¹⁶ Using Cinestream 3.1, we conducted a frame-by-frame analysis of the video footage to determine each infant's gaze direction and duration. Gaze duration was summed for each video image (ie, the /a/ articulation and the /i/ articulation) and averaged across stimulus conditions. This yielded the mean total looking time (in seconds) for each image during the test phase. These analyses resulted in a minimum of data from 1

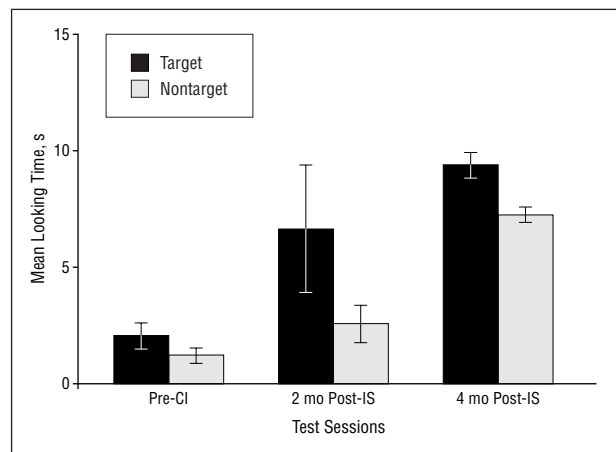


Figure 3. Individual bimodal perception data from infant cochlear implant (CI) user CI-01, who was 14 months 8 days old at initial stimulation (IS). Error bars indicate standard error.

preimplant hearing aid test session and 1 post-CI stimulation test session for each participant.

STATISTICAL ANALYSIS

Descriptive statistics were used to describe the basic features of the preliminary data from this study. Paired *t* tests were used to compare the mean looking times across test sessions for each infant and across test sessions for the group.

RESULTS

In the present experiment, a proper matching of the auditory and visual stimuli was assumed if the infant spent significantly more time looking at the woman articulating the same vowel sound presented via a loudspeaker. Examples of some individuals' SPLP data, followed by preliminary group data, are presented herein.

Patient CI-01 was a consistent bilateral hearing aid user prior to implantation; his results are displayed in **Figure 3**. Although none of the differences in mean looking time were statistically significant, his data illustrate a clear developmental trend. During his preoperative hear-

ing aid test session, his looking times were low, and there was no difference between his gaze directions ($t_3=0.93$, $P>.05$). By the second month after stimulation, his looking times increased and began to show a trend in the predicted direction ($t_3=0.62$, $P>.05$). Finally, by his most recent test session, his looking times continued to increase, and he spent more time looking at the target video that correctly matched the vowel presented via the loudspeaker, although it did not reach statistical significance ($t_3=0.26$, $P>.05$). After 4 months of CI experience, it appears he was beginning to match the vowel information from the talker's lips and voice.

Figure 4 displays data from implant user CI-02, who was also a consistent bilateral hearing aid user before surgery. There was great variability in her task performance across test sessions and no clear developmental trend. Although this infant spent much of her time looking at both of the video stimuli during her first 3 test sessions, she did not spend significantly more time looking at a particular face when the appropriate vowel sound was heard (pre-CI, $t_3=-0.99$, $P>.05$; 2 months after initial CI stimulation [IS], $t_3=0.30$, $P>.05$; and 4 months post-IS, $t_3=-0.14$, $P>.05$). During her 9-month test session, however, she appeared to have mastered the task and spent significantly more time looking at the target video ($t_3=3.01$, $P<.05$). These data suggest that patient CI-02 was able to successfully match the phonetic information in this task by 9 months post-IS.

Finally, **Figure 5** shows data from patient CI-03, a child whose performance across test sessions is contrary to anything predicted. During the initial hearing aid test session, this infant was already able to match the vowel information from the talker's lips and voice ($t_3=4.62$, $P<.01$), yet after 2 months of CI experience, his behavior in the bimodal task changed significantly, and he no longer demonstrated the ability to correctly match the video image and vowel presented via the loudspeaker (2 months post-IS, $t_3=1.94$, $P>.05$; 4 months post-IS, $t_3=-0.76$, $P>.05$; and 6 months post-IS, $t_3=-1.81$, $P>.05$). A retrospective examination of factors that might have contributed to this change indicated that the patient's comfort levels were programmed quite high during his 2-month post-IS test session. It is possible that the differences in his looking times across test sessions were a result of a thorough hearing aid fitting before CI implantation and/or the challenges associated with determining the ideal parameters that define the processing of audio signals into digital signals (ie, the MAP) for a child of this age.

Preliminary group data from the infant CI users is displayed in **Figure 6**. This figure also shows that despite the large amount of variability in this small data set there appears to be a developmental trend across test sessions. However, mean looking times to the target and nontarget stimuli did not reach statistical significance until the 9-month post-IS test session (pre-CI, $t_6=-1.33$, $P>.05$; 2-month post-IS, $t_3=1.56$, $P>.05$; 4-month post-IS, $t_6=1.18$, $P>.05$; 6-month post-IS, $t_2=0.54$, $P>.05$; and 9-month post-IS, $t_3=3.01$, $P<.05$). As the infants' levels of audibility improved and they gained more listening experience with their CIs, their mean target looking times increased across test sessions. This suggests that the group's improved levels of audibility subsequently may

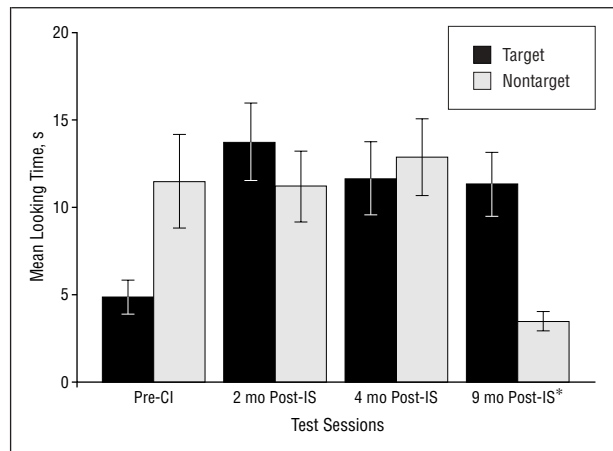


Figure 4. Individual bimodal perception data from infant cochlear implant (CI) user CI-02, who was 13 months, 3 days old at initial stimulation (IS). Error bars indicate standard error; asterisk, $P<.05$.

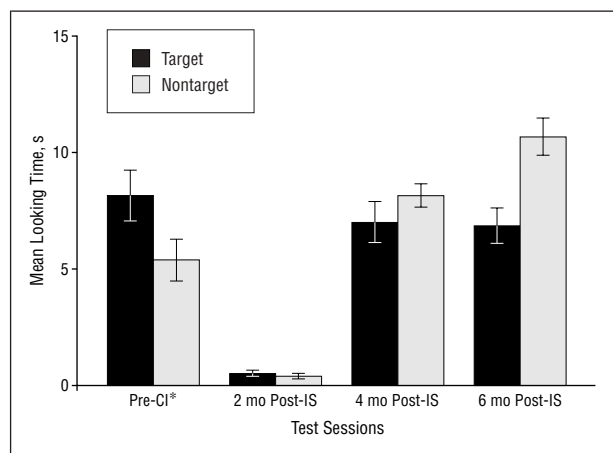


Figure 5. Individual bimodal perception data from infant cochlear implant (CI) user CI-03, who was 14 months 12 days old at initial stimulation (IS). Error bars indicate standard error; asterisk, $P<.05$.

have contributed to their ability to begin matching phonetic information in the lips and voice for the present task.

COMMENT

The preliminary results from this bimodal perception experiment are promising. They demonstrate that the results of earlier bimodal perception studies are reproducible^{9,12} and thus offer the hope that such research can be used to develop an empirically valid clinical protocol to assess the auditory benefit of hearing instruments for hearing-impaired infants.

These preliminary data, however, do not allow one to draw any solid conclusions regarding the exact point in development at which these infants are able to match phonetic information from the lips and voice. To determine this point in development, current efforts are focused on recruiting more participants to increase statistical power and establish complete longitudinal data sets. Speech perception phenomena to date have been evaluated using cross-sectional designs, and these data have been combined to form a "complete" picture of cognitive development. The lack of longitudinal data in normal-hearing infants means

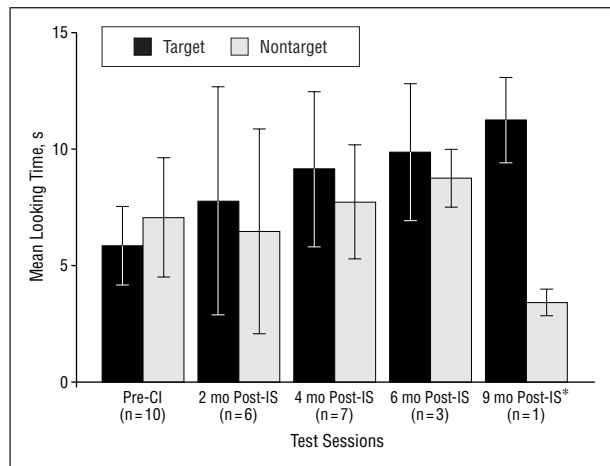


Figure 6. Group bimodal perception data from infant implant users gathered before and after placement of the cochlear implant (CI). Error bars indicate standard error; IS, initial CI stimulation; asterisk, $P < .05$.

that we do not know how infants continue to perform on speech perception tasks after there is evidence of reaching particular performance levels. Collection of longitudinal data for both normal-hearing infants and CI users will ultimately reveal the shape of the growth function. If the function is not linear (as it is often assumed to be) but rather an inverted U shape, one may actually miss the developmental zenith of performance in the older participants (eg, the CI users). For this reason, additional data must be gathered before reliable conclusions can be made regarding infant CI users' bimodal perception skills.

Nonetheless, the present results suggest that the auditory experience provided to the infants via their CIs may have improved their levels of audibility and thus laid the foundation for bimodal speech perception. (It is worth noting that the individual data suggest that for some children [eg, CI-03], even the auditory input provided via hearing aids may be sufficient for the bimodal perception of the vowels /a/ and /i/.)

In conclusion, the present experiment prompts numerous directions for future research. First, the auditory input provided via a CI may be less than optimal, given the challenges associated with creating an ideal MAP for an infant (eg, for some infants one may be unable to obtain electrophysiologic measures or behavioral measures to aid in programming the CI). At a glance, it appears that these challenges were reflected in the differences noted in audibility and comfort levels for participant CI-03 in the present experiment. It would be valuable to examine these differences in detail and begin to explore possible correlations between audibility, comfort levels, and success with the various speech perception tasks. It would also be fruitful to compare the emergence of bimodal perception and canonical babbling in these children as a means of exemplifying the speech perception–production link noted in older pediatric CI users.¹⁰ Finally, it is crucial that additional data be gathered from infants with a variety of hearing levels so that we can begin to better understand the

development of speech perception skills across infant hearing aid users as well as infant CI users.

Submitted for publication September 10, 2003; accepted December 9, 2003.

This work was supported in part by research grant 2 P50 DC00242 from the National Institute on Deafness and Other Communication Disorders and grant RR00059 from the General Clinical Research Centers Program, Division of Research Resources, National Institutes of Health, Bethesda, Md; the Lions Clubs International Foundation, Oak Brook, Ill; and the Iowa Lions Foundation, Ames.

This study was presented at the Ninth Symposium on Cochlear Implants in Children; April 25, 2003; Washington, DC.

We would like to thank Victoria C. Klein, BA, for engaging in extensive coding of the infants' test sessions. We also thank Courtney M. Burke, MA, and Michelle L. Hughes, PhD, for their help with the stimuli creation and Linda J. Spencer, MA, and Sandie M. Bass-Ringdahl, PhD, for their helpful comments during this study. Finally, we would like to express our gratitude to the infants and families who have volunteered so much of their time to the Children's Cochlear Implant Program at the University of Iowa.

Corresponding author and reprints: Brittan A. Barker, MA, Department of Otolaryngology–Head and Neck Surgery, University of Iowa Hospitals and Clinics, 200 Hawkins Dr, 21200 PFP, Iowa City, IA 52242 (e-mail: brittan-barker@uiowa.edu).

REFERENCES

- DeCasper AJ, Fifer WP. Of human bonding: newborns prefer their mothers' voices. *Science*. 1980;208:1174-1176.
- DeCasper AJ, Lecanuet J-P, Busnel M-C, Granier-Deferre C. Fetal reactions to recurrent maternal speech. *Infant Behav Dev*. 1994;17:159-164.
- DeCasper AJ, Spence MJ. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav Dev*. 1986;9:133-150.
- Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. *Science*. 1982;218:1138-1140.
- Kuhl PK, Meltzoff AN. The bimodal representation of speech infants. *Infant Behav Dev*. 1984;7:361-381.
- Kuhl PK, Williams KA, Meltzoff AN. Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *J Exp Psychol Hum Percept Perform*. 1991;17:829-840.
- Legerstee M. Infants use multimodal information to imitate speech sounds. *Infant Behav Dev*. 1990;13:343-354.
- Walton GE, Bower TGR. Amodal representation of speech in infants. *Infant Behav Dev*. 1993;16:233-243.
- Patterson ML, Werker JF. Two-month-old infants match phonetic information in lips and voice. *Dev Sci*. 2003;6:191-196.
- Lachs L, Pisoni DB, Kirk KI. Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: a first report. *Ear Hear*. 2001;22:236-250.
- Osberger MJ, Miyamoto RT, Zimmerman-Phillips S, et al. Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear Hear*. 1991;12:66S-80S.
- Patterson ML, Werker JF. Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behav Dev*. 1999;22:237-247.
- Bayley N. *Bayley Scales of Infant Development II*. San Antonio, Tex: Psychological Corp; 1993.
- Patterson ML, Werker JF. Infants' ability to match dynamic phonetic and gender information in the face and voice. *J Exp Child Psychol*. 2002;81:93-115.
- Golinkoff RM, Chung HL, Hirsh-Pasek K, et al. Young children can extend motion verbs to point-light displays. *Dev Psychol*. 2002;38:604-614.
- Hollich GJ. Guide to the Splitscreen Preferential Looking Paradigm. 2003. Available at: <http://hincapie.psych.purdue.edu/Splitscreen/index.html>. Accessed January 20, 2004.