DRDC Toronto CR 2005-064

SPEECH INPUT HARDWARE INVESTIGATION FOR FUTURE DISMOUNTED SOLDIER COMPUTER SYSTEMS

by:
Jeffrey C. Bos and David W. Tack

Human*systems*® Incorporated
111 Farquhar St., 2nd floor
Guelph, ON N1H 3N4

Project Manager:
David W. Tack
(519) 836 5911

PWGSC Contract No. W7711-017747/001/TOR
Call-Up 7747-24
HSI® SIREQ Item #152

On behalf of
DEPARTMENT OF NATIONAL DEFENCE

as represented by
Defence Research and Development Canada - Toronto
1133 Sheppard Avenue West
Toronto, Ontario, Canada
M3M 3B9

DRDC Toronto Scientific Authority
LCol Linda Bossi
(416) 635-2197

May 2005

Abstract

Speech input devices are required by future infantry solders to provide voice input to radio communications systems for voice communication between individuals and to provide speech input to electronic speech recognition systems. A comparison of six types of audio input devices (bone conduction, in-ear, boom, throat and physiological microphones and lip tracking continuous speech recognition) was conducted based upon existing literature, user feedback and the authors' experience. This report presents the results of that comparison. The results indicate some devices are more compatible with certain soldiering tasks. Overall, the bone conduction microphone is rated better than other types of speech input devices. While the bone conduction microphone has no unique advantages over the other devices in terms of Speech Input Performance, Equipment Compatibility, Fit and Comfort, Tactical Feasibility, Reliability, and Availability and Cost, it achieves consistently high ratings in terms of Free Field Sound Detection, Noise Attenuation, and Durability. The physiological microphone was rated second most acceptable overall. The physiological microphone has unique advantages over all other headsets in terms of Equipment Compatibility and Fit and Comfort. The physiological microphone would likely achieve the highest overall rating if a reliable COTS (Commercial Off-The-Shelf) version was available with acceptable audio quality. The boom microphone was rated third most acceptable overall. The boom microphone also has unique advantages in terms of Speech Input Performance, Reliability, and Availability and Cost. The boom microphone is the most familiar and conventional headset. It is durable and accurate but is very susceptible to free field noise.

Résumé

Dans l'avenir, les fantassins auront besoin de dispositifs de reproduction vocale pour fournir des signaux vocaux aux systèmes de radiocommunications utilisés pour les transmissions vocales entre personnes ainsi qu'à des circuits électroniques de reconnaissance de la parole. Une comparaison de six dispositifs de reproduction audio (microphone à conduction osseuse, microphone auriculaire, micro-rail, laryngophone, microphone physiologique et dispositif de reconnaissance de la parole continue par suivi des lèvres) a été faite compte tenu de la documentation existante, la rétroaction des utilisateurs et l'expérience des auteurs. Les résultats indiquent que des dispositifs particuliers conviennent mieux à certaines tâches de la vie de soldat. Dans l'ensemble, le microphone à conduction osseuse s'est classé avant les autres types de dispositifs d'entrée vocale. Bien que le microphone à conduction osseuse n'offre aucun avantage unique du point de vue du rendement avec entrée vocale, de la compatibilité de l'équipement, de l'adaptation et du confort, de la faisabilité tactique, de la fiabilité ainsi que de la disponibilité et des coûts, il s'est constamment classé parmi les meilleurs pour ce qui est de la détection sonore en champ libre, de l'atténuation du bruit et de la durabilité. Le microphone physiologique s'est classé au deuxième rang. Il offre des avantages uniques pour ce qui est de la compatibilité de l'équipement ainsi que de l'adaptation et du confort. Si un microphone physiologique fiable disponible sur le marché offrait une qualité audio acceptable, il obtiendrait vraisemblablement la meilleure note globale. Le micro-rail est arrivé en troisième position. Il offre aussi des avantages uniques quant au rendement avec entrée vocale, à la fiabilité ainsi qu'à la disponibilité et aux coûts. Il s'agit du dispositif le plus répandu et le plus ordinaire. Durable et précis, il est toutefois très vulnérable au bruit en champ libre.

Executive Summary

Systems are required to provide voice input to radio communications systems for voice communication between individuals and to provide speech input to electronic speech recognition systems. A comparison of six types of audio input devices was conducted based upon existing literature, user feedback and the authors' experience. This report presents the results of that comparison. Options investigated include Boom Microphone, Throat Microphone, In-Ear Microphone, Bone Conduction Microphone, Physiological Microphone, and Lip Tracking Continuous Speech Recognition.

The devices were evaluated and compared according to nine performance criteria: Free Field Sound Detection, Speech Input Performance, Equipment Compatibility, Fit & Comfort, Noise Attenuation, Tactical Feasibility, Durability, Reliability, and Availability and Cost. A list of advantages and disadvantages was created for each type of device based on review of the available product information, experience evaluating the devices during previous field trials, and expert human factors evaluation based on prior domain knowledge and user feedback. Weights were systematically assigned to each of the performance rating criteria.

Each system has special features that could provide advantages to soldiers. The Boom Microphone is a very reliable system that provides excellent signal capture and is relatively inexpensive. The Throat Microphone is highly resistant to ambient noise and is able to detect low volume speech. The In-Ear Microphone is also able to detect low volume speech and is available in a wireless configuration. It can be used to transmit and receive sound. The Bone Conduction Microphone is resistant to ambient noise and can be integrated with helmets. It can also be used to both transmit and receive sound. The Physiological Microphone is highly resistant to ambient noise and can serve as a health monitor. The Lip Tracking Continuous Speech Recognition system can detect words tracking lip movements in complete stealth.

Several disadvantages were identified for each of the devices. The Boom Microphone is highly susceptible to ambient noise and would require an electronic noise cancellation system for use in loud operational environments. The Throat Microphone is problematic because it must stay firmly in contact with a specific area of the wearer's neck. If the strap stretches or becomes loose, the microphone can not detect a signal. The In-Ear Microphone sits inside the wearer's ear and could potentially cause ear infections if not properly maintained. The Bone Conduction Microphone must also stay in direct contact with the wearer in order to transmit a signal. The Physiological Microphone must stay in contact with the wearer and has not yet been proven as a reliable technology. The Lip Tracking Continuous Speech Recognition system is only in the early stages of development.

Performance ratings were tallied and devices were ranked according to their total score. Overall, the bone conduction microphone was rated better than other types of speech input devices. While the bone conduction microphone has no unique advantages over other devices in terms of Speech Input Performance, Equipment Compatibility, Fit and Comfort, Tactical Feasibility, Reliability, and Availability and Cost, it achieved consistently high ratings in terms of Free Field Sound Detection, Noise Attenuation, and Durability.

The Physiological Microphone was rated second most acceptable overall. It has unique advantages over all other headsets in terms of Equipment Compatibility and Fit and Comfort. If a reliable COTS (Commercial Off-The-Shelf) version of the Physiological Microphone were available with superior audio quality, it would likely achieve the highest overall rating.

The Boom Microphone was rated third most acceptable overall. The Boom Microphone also has unique advantages in terms of Speech Input Performance, Reliability, and Availability and Cost. The Boom is the most familiar and conventional type of headset. It is durable and accurate but is susceptible to free field sounds.

The Bone Conduction, Physiological, and Lip Tracking input devices were all rated best in terms of Free Field Noise Attenuation. The Bone Conduction and Physiological Microphones are rated better than the other display types in terms of Durability. The In-Ear Microphone was the only headset that did not receive the highest possible acceptability rating for Free Field Sound Detection because it fully occludes one of the wearer's ears.

Sommaire

Des dispositifs sont nécessaires pour fournir des signaux vocaux aux systèmes de radiocommunications utilisés pour les transmissions vocales entre personnes ainsi qu’à des circuits électroniques de reconnaissance de la parole. Six dispositifs de reproduction audio ont été comparés compte tenu de la documentation existante, de la rétroaction des utilisateurs et de l’expérience des auteurs. Les résultats sont présentés dans ce rapport. Voici les dispositifs qui ont été comparés : micro-rail, laryngophone, microphone auriculaire, microphone à conduction osseuse, microphone physiologique et dispositif de reconnaissance de la parole continue par suivi des lèvres.

Les dispositifs ont été évalués et comparés selon neuf critères de rendement : détection sonore en champ libre, rendement avec entrée vocale, compatibilité de l’équipement, adaptation et confort, atténuation du bruit, faisabilité tactique, durabilité, fiabilité, et disponibilité et coûts. Une liste d’avantages et d’inconvénients a été créée pour chaque type de dispositifs, selon l’examen de l’information disponible sur les produits, l’expérience tirée de l’évaluation de dispositifs durant des essais antérieurs sur le terrain et l’évaluation de facteurs humains par des experts à partir de la connaissance antérieure du domaine et des réactions des utilisateurs. Une pondération a été affectée systématiquement à chacun des critères d’évaluation du rendement.

Chaque dispositif offre des caractéristiques spéciales qui pourraient se révéler avantageuses pour les fantassins. Le micro-rail est un dispositif très fiable qui capte très bien les signaux et n’est que relativement peu coûteux. Le laryngophone est très résistant au bruit ambiant et peut détecter la parole à bas volume. Le microphone auriculaire peut également détecter la parole à bas volume et existe dans une configuration sans fil. Elle peut servir à émettre et à recevoir les sons. Le microphone à conduction osseuse est résistant au bruit ambiant et peut s’intégrer à un casque. Il peut également servir à émettre et à recevoir les sons. Le microphone physiologique est très résistant au bruit ambiant et peut également servir au contrôle d’état physiologique. Le système de reconnaissance de la parole continue par suivi des lèvres peut détecter furtivement les mots prononcés en suivant le mouvement des lèvres.

Plusieurs inconvénients ont été mis en évidence pour chacun de ces dispositifs. Le micro-rail est très sensible au bruit ambiant et nécessiterait un système électronique d’annulation du bruit pour les environnements opérationnels particulièrement bruyants. Le laryngophone pose problème du fait qu’il doit rester fermement en contact avec une zone déterminée de la gorge de l’utilisateur. Si la courroie s’étire ou se relâche, le microphone n’est plus capable de détecter les signaux. Le microphone auriculaire reste à l’intérieur de l’oreille de l’utilisateur et peut occasionner des infections si elle n’est pas correctement entretenue. Le microphone à conduction osseuse doit également demeurer en contact direct avec l’utilisateur pour l’émission d’un signal. Le microphone physiologique doit rester en contact avec l’utilisateur, et la technologie sur laquelle il repose n’a pas encore fait ses preuves. Le système de reconnaissance de la parole continue par suivi des lèvres n’en est qu’à ses premiers stades de développement.

Des évaluations de rendement ont été établies, et les dispositifs ont été classés en fonction de leur note totale. Dans l’ensemble, le microphone à conduction osseuse s’est classé avant les autres types de dispositifs d’entrée vocale. Bien que le microphone à conduction osseuse n’offre aucun avantage unique du point de vue du rendement avec entrée vocale, de la compatibilité de l’équipement, de l’adaptation et du confort, de la faisabilité tactique, de la fiabilité ainsi que de la disponibilité et des

coûts, il s’est constamment classé parmi les meilleurs pour ce qui est de la détection sonore en champ libre, de l’atténuation du bruit et de la durabilité.

Le microphone physiologique s’est classé au deuxième rang. Il offre des avantages uniques pour ce qui est de la compatibilité de l’équipement ainsi que de l’adaptation et du confort. Si un microphone physiologique fiable disponible sur le marché offrait une qualité audio supérieure, il obtiendrait vraisemblablement la meilleure note globale.

Le micro-rail est arrivé en troisième position. Il offre aussi des avantages uniques quant au rendement avec entrée vocale, à la fiabilité ainsi qu’à la disponibilité et aux coûts. Il s’agit du dispositif le plus répandu et le plus ordinaire. Durable et précis, il est toutefois vulnérable au bruit en champ libre.

Le microphone à conduction osseuse, le microphone physiologique et le dispositif de reconnaissance de la parole continue par suivi des lèvres se sont tous classés au meilleur niveau pour ce qui est de l’atténuation du bruit en champ libre. Les microphones à conduction osseuse et physiologiques devancent les autres types du point de vue de la durabilité. Le microphone auriculaire est le seul dispositif à ne pas avoir atteint la plus haute note d’acceptabilité du point de vue de la détection sonore en champ libre, car elle bouche complètement l’une des oreilles de l’utilisateur.

Table of Contents

List of Tables

List of Figures

1. Introduction

An investigation was conducted to compare audio input devices required by future infantry solders to provide an input capability for maintaining radio communications between individuals and for interacting with electronic speech recognition systems. As a basis for distinguishing between the different types of systems, the taxonomy of speech input devices was generated by compiling a comprehensive list of existing, commercially available, and experimental voice input devices. Speech input devices were classified according to: Microphone Type, Fit, and Noise Reduction. Based on this taxonomy, six types of speech input devices were investigated: Boom Microphone, Throat Microphone, In-Ear Microphone, Bone Conduction Microphone, Physiological Microphone, and Lip Tracking Continuous Speech Recognition (see Table 1).

In some configurations, a speech input device may partially or entirely cover the wearer's mouth. In cases of full helmet encapsulation, it must be possible for the wearer to speak outside of the system or perhaps generate a local broadcast using a public address speaker system.. The speech-input system is required to have sufficient sensitivity and clarity to enable whispered communication while guaranteeing the intelligibility of speech in a noisy environment. The system is also required to minimize the risk of snagging. The criteria used to evaluate each system are described below in the next section.

Table 1. Speech Input Hardware Taxonomy

Input Device Attributes	Boom Microphone	Throat Microphone	In-Ear Microphone	Bone Conduction Microphone	Physiological Microphone	Lip Tracking Continuous Speech Recognition
Microphone Type	Open Air	Vibration Sensor	Intra-aural	Attached to Head Band (or helmet mounted)	Attached to Head Band (or helmet mounted)	Visual pattern detector
Fit	Microphone mounted on Boom attached to Helmet or Head Band	Neck Band	Universal (or custom) fit Plastic or Foam	Must remain in contact with jaw or scull	Must remain in contact with soft tissue	Sensor mounted on Boom attached to Helmet or Head Band
Noise Reduction	No (or Active or Passive)	No (or Passive or Active)	No (or Passive or Active)	Passive (or Active)	Passive (or Active)	N/A
Features		Detects low volume speech	Wireless or wired	Resistant to ambient noise	Resistant to ambient noise, can be used w/health monitor, & water-proof	Resistant to ambient noise

2. Method

Prior to conducting this investigation, the authors acquired extensive domain knowledge participating in several field trials to evaluate various audio headset configurations as part of the SIREQ TD program. These field trials evaluated the Marconi PRR headset and various alternative headset configurations (Woods, Tack, and Adams, 2003) for radio communication and for electronic speech recognition (Bos and Tack, 2004).

In order to evaluate and compare these systems, a set of performance rating criteria were defined considering relevant measures of human performance and the context of use (see Table 2). A list of advantages and disadvantages was created for each type of device based on review of the available product information, experience evaluating the devices during previous field trials, and expert human factors evaluation based on prior domain knowledge and user feedback.

Table 2. Definitions of Rating Criteria for Speech Input Devices

Rating Criteria	Definition of Rating Criteria
Free Field Sound Detection	Some types of speech input devices are worn over or inside of the ear (e.g. bone conduction microphones). Systems must not interfere with the wearer's ability to detect sounds and hear voices in the free field while wearing the speech input device.
Speech Input Performance	Clarity, loudness, and intelligibility of speech transmitted by the speech-input device in ideal conditions.
Equipment Compatibility	Compatibility of the speech input device with other types of equipment or clothing.
Fit & Comfort	Security of fit on the wearer and degree of comfort when wearing the speech input device.
Noise Attenuation.	The microphone's ability to transmit speech while attenuating background noise in the free field
Tactical Feasibility	Ability to move tactically without detection while wearing and speaking into the speech input device.
Durability:	A measure of the strength of construction of the speech input device and the device's ability to endure field use without breaking.
Reliability	An estimate of the wearer's ability to repeatedly transmit speech with the device without errors.
Availability and Cost	The relative price and commercial availability of the speech input device.

2.1 Boom Microphones

Boom microphones use an open, unidirectional transducer and are positioned in front of the wearer's mouth. They provide passive noise attenuation and can be used in conjunction with active noise reduction systems. Noise in the open field can interfere with the signal quality of open microphones.

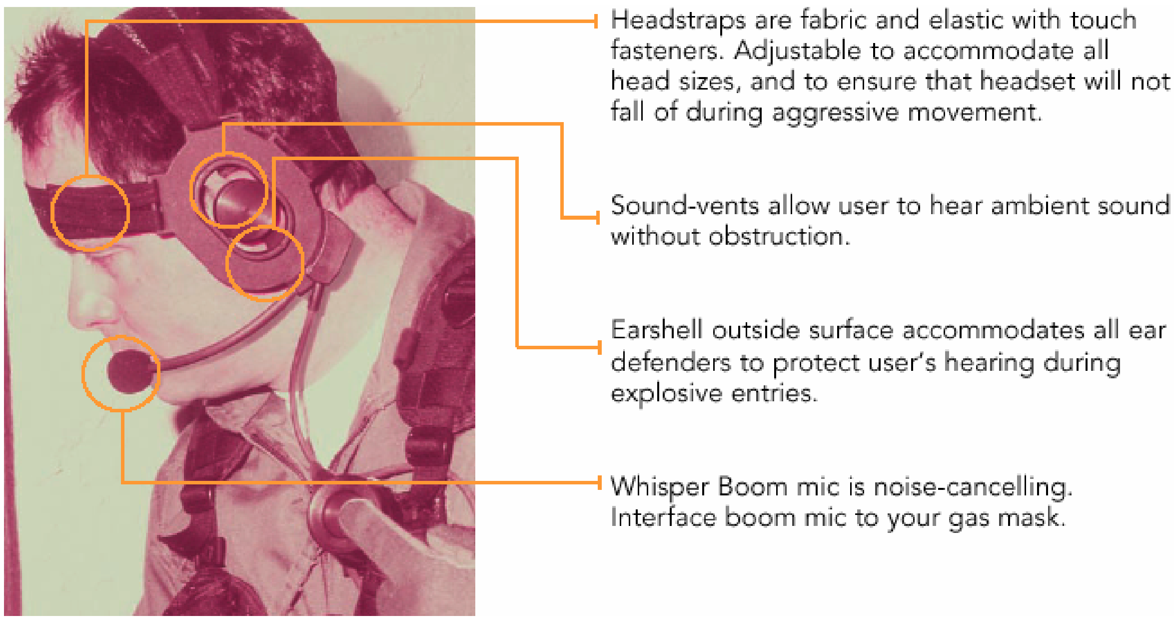

Figure 1. Lite2 Boom Microphone

2.1.1 Advantages

a) Free Field Sound Detection: The boom microphone component of the headset does not cover the ear or obstruct the wearer's hearing in any way. It does not interfere with the wearer's ability to detect sounds in the free field.

b) Speech Input Performance: Boom microphones can provide optimal sound quality for transmitting voice communications. Boom microphones are used for professional audio recording applications.

c) Equipment Compatibility: The boom can be either attached to a helmet or to a head mount worn beneath a helmet.

d) Fit/Comfort: The microphone is positioned in front of the wearer's mouth when it is in use. The flexible boom can be moved away from the wearer's face when it is not in use.

e) Tactical Feasibility: Boom microphones can detect whispered and low volume speech. Soldiers can cup their hand over the microphone and whisper to avoid enemy detection.

f) Noise Attenuation: Unidirectional microphones (e.g. Cardioid microphones) use directional sound input to reduce interference from unwanted external sound sources

g) Reliability: Boom microphones are mechanically simple and reliable.

h) Durability: The microphone can be built to withstand shock and impact damage.

i) Availability and Cost: Boom microphones are inexpensive to manufacture and are commercially available.

2.1.2 Disadvantages

a) Equipment Compatibility: The boom can interfere with the soldier's helmet or rifle.

b) Fit/Comfort: The boom can brush against the wearer's face. Soldiers must adjust the boom to ensure the microphone is positioned in front of the mouth when transmitting. Vigorous physical activity can dislodge or misalign the microphone.

c) Noise Attenuation: All types of acoustic boom microphones are sensitive to background noise. Wind blowing on the microphone can create a low rumble that interferes with speech intelligibility.

2.2 Throat Microphones

Throat microphones use a vibration sensor that must remain in continuous contact with the wearer's throat. They provide passive noise attenuation and can be used in conjunction with active noise reduction. Throat microphones are not affected by noise in the open field.

Figure 2. Thales RA440 Throat Microphone

2.2.1 Advantages

a) Free Field Sound Detection: Throat microphones do not interfere with the wearer's ability to detect sounds in the free field.

b) Speech Input Performance: The throat microphone transmits a louder voice than other microphones because of the microphone's proximity to the sound source.

c) Noise Attenuation: Throat microphones are designed to be resistant to ambient noise in the free field.

d) Equipment Compatibility: The throat microphone neckband can be worn inconspicuously beneath the wearer's clothing.

e) Fit/Comfort: The throat microphone neckband provides comfortable, inconspicuous, and stable positioning when properly adjusted.

f) Tactical Feasibility: The throat microphone can detect whispered speech for tactical use.

g) Noise Attenuation: Throat microphones are resistant to background noise in the free field. The wearer's voice can be easily heard over other voices and sounds in the background.

h) Reliability: The neckband attaches around the soldier's neck to secure the microphone in place at all times during physical activity.

i) Availability and Cost: Throat microphones are inexpensive to manufacture and are commercially available.

2.2.2 Disadvantages

a) Speech Input Performance: The wearer's voice will not be as clear as with a microphone at the lip line, because this unit will amplify body noises, (i.e. when turning the head rapidly, and swallowing, any extraneous sounds). The wearer's speech lacks lip articulation because speech is transmitted at the voice box rather than at the lip line.

b) Fit/Comfort: The microphone must be positioned in the wearer's "sweet spot" in order to transmit speech. The microphone will not transmit speech if it is positioned over the wearer's thyroid cartilage (Adam's Apple). The neckband can become restrictive and constrict breathing if it is attached incorrectly or adjusted too tight. Dirt and debris can get caught between the wearer's neck and the neckband causing skin irritation and discomfort.

c) Reliability: The neckband tension adjuster can slip or the neckband can stretch with prolonged use, causing the microphone to become slack and fail to detect speech.

d) Durability: The neckband must be constructed of elastic material that is susceptible to stretching or tearing. Moisture and dirt can accumulate on the neck strap and cause the fibres of the material to deteriorate and lose it's elasticity.

2.3 In-Ear Microphone

In-ear microphones use either a vibration sensor that must remain in continuous contact with soft tissue in the wearer's ear canal. They provide passive noise attenuation and can be used in conjunction with active noise reduction. In-ear microphones are not affected by noise in the open field.

Figure 3. Invisio Ear Microphone

2.3.1 Advantages

a) Speech Input Performance: Speech quality is very good because the microphone detects speech transmitted within the ear canal.

b) Noise Attenuation: The in-ear microphone provides excellent sound insulation from external noises in the free field.

c) Equipment Compatibility: In-ear microphones are compatible with ear defenders or with circumaural headsets. In-ear microphones can provide active or passive talk through and can be used with active hearing enhancement devices.

d) Fit/Comfort: In-ear microphones are designed to fit snugly inside the ear canal. Custom earmolds can be made to fit the individual wearer. The earplug provides comfortable, inconspicuous, and stable positioning when properly inserted.

e) Tactical Feasibility: The in-ear microphone can detect whispered speech for tactical use.

f) Reliability: The in-ear microphone is firmly secured in place to reduce the probability of becoming misaligned to the ear as a result of vibration or physical activity.

g) Availability and Cost: In-ear microphones are commercially available.

2.3.2 Disadvantages

a) Free Field Sound Detection: In-ear microphones interfere with the soldier's ability to detect free-field sounds because the enclosure prevents external sounds from reaching the ear. Passive or active talk-through systems must be used to provide access to external sounds. Talk-through systems are not as accurate as the unimpeded ear.

b) Fit/Comfort: In ear microphones cause discomfort with prolonged use because they can irritate the skin in the ear canal. Hygiene is a concern with any device worn inside the ear. Dirt or other debris that enters the ear canal can cause ear lacerations or inner ear infections.

c) Reliability: In-ear microphones must be positioned in the wearer's "sweet spot" to transmit audible speech. Wires connected to the ear microphone can pull on the wearer's ear and cause the microphone to lose contact with the "sweet spot" or fall out.

d) Durability: When not worn inside the ear, the earplug is a fragile device that can be easily contaminated or damaged and rendered unserviceable.

e) Availability and Cost: In-ear microphones are commercially available but are expensive to manufacture. Custom earmolds are more expensive to manufacture than universal earplugs because an audiologist must take measurements of the inner ear.

2.4 Bone Conduction Microphone

Bone conduction microphones use a vibration sensor that must remain in continuous contact with the wearer's head. They provide no inherent noise attenuation but can be used in conjunction with passive or active noise reduction devices. Bone conduction microphones are not affected by noise in the open field.

Figure 4. Temco Voiceducer Headgear HG17

2.4.1 Advantages

a) Free Field Sound Detection: Bone conduction headsets do not interfere with the detection of sounds or verbal communications in the free field. The headset fits away from the ear, so the wearer's ears are clear to listen to sounds in the free field.

b) Speech Input Performance: Speech quality is very good because the microphone detects speech transmitted through the bones in the skull.

c) Noise Attenuation: Bone conduction microphones are resistant to background noise in the free field. The wearer's voice can be easily heard over other voices and sounds in the background.

d) Equipment Compatibility: The microphone can be integrated with the wearer's helmet for a precise fit. Bone conduction microphones are also compatible with ear defenders and with circumaural headsets. Bone conduction microphones do not cover the ears and provide open talk through. They are compatible with active noise reduction and active hearing enhancement.

e) Fit/Comfort: Bone conduction microphones are small and light enough to be worn for an extended period of time. These microphones are not obtrusive to the wearer because they are worn away from the face. Bone conduction microphones provide comfortable, inconspicuous, and stable positioning when secured correctly.

f) Tactical Feasibility: Bone conduction microphones can detect whispered speech for tactical use. The microphone can be hidden beneath a hat or helmet for inconspicuous speech communications.

g) Reliability: The bone conduction microphone can be firmly secured in place by the wearer's helmet to reduce the probability of it becoming misaligned to the ear as a result of vibration or physical activity. Cranial bone conduction microphones are less sensitive to correct positioning than jaw mounted or in-ear bone conduction microphones.

h) Durability: The bone conduction transducer is constructed of metal. The headset can withstand vibration and shock without failure. but

i) Availability and Cost: Bone conduction microphones are commercially available.

2.4.2 Disadvantages

a) Equipment Compatibility: Some helmet types are not compatible with some types of bone conduction microphones. Incompatible helmets do not provide sufficient clearance to enable the microphone to fit beneath the helmet. The microphone can create a pressure point on the wearer's scalp if it does not fit properly beneath a helmet.

b) Reliability: The bone conduction microphone must be firmly in contact with the wearer's head in order to transmit speech. If contact is interrupted or lost, no speech is transmitted.

c) Availability and Cost: Ruggedized bone conduction microphones are expensive to manufacture.

2.5 Physiological Microphone

United States Army Research Lab has developed sensor technology to monitor a wearer's voice and physiology by using enhanced acoustic sensors. This sensor detects speech by measuring acoustic signals through the speaker's skin. While the signal produced is not typical of that from an airborne acoustic microphone, the possibility exists for using this sensor as a microphone.

The sensors consist of a fluid or gel contained within a small, conformable, rubber bladder or pad that also includes a hydrophone. This enables the collection of high signal-to-noise ratio (SNR) cardiac, respiratory, voice, and other physiological data. The pad also minimizes interference from ambient noise because it couples poorly with airborne noise. It is low cost and comfortable to wear for extended periods.

ARL further developed (with support from the United States Military Academy), a Hidden-Markov Model (HMM) phoneme-level speech system optimized for the physiological sensor package. Lab observations indicated tremendous potential for use in automatic speech recognition under quiet and very high noise conditions. The next step involves development of an experimental design to test the effectiveness of the sensor in a more controlled setting.

Figure 5. Physiological Microphone Neckstrap (top left), Goggle Mount (top right), Neckstrap Detail (bottom left), and Goggle Mount Detail (bottom left).

2.5.1 Advantages

a) Free Field Sound Detection: The physiological microphone does not interfere with the wearer's ability to detect sounds in the free field.

b) Speech Input Performance: Audible speech can be transmitted through bone and soft tissues of the head.

c) Noise Attenuation: Physiological microphones are resistant to background noise in the free field. The wearer's voice can be easily heard over other voices and sounds in the background.

d) Equipment Compatibility: The physiological microphone is designed to be in direct contact with the body. It can be integrated with the wearer's helmet, worn on the wrist, torso, or any other part of the body.

e) Fit/Comfort: The soft gel pads provide a universal fit and can absorb impact and vibration during movement.

f) Tactical Feasibility: Physiological microphones can detect whispered speech for tactical use. The microphone can be hidden beneath a hat, helmet or clothing for inconspicuous speech communications. It can also be used to detect the wearer's health, fatigue, and posture.

g) Reliability: The microphone can be tightly secured to the wearer without interfering with mobility or freedom of movement.

h) Durability: The flexible gel pads resist shock, impact, and puncture damage. The gel pads are resistant to moisture and heat.

i) Availability and Cost: Physiological microphones are commercially available from a company called RadioEar.

2.5.2 Disadvantages

a) Speech Input Performance: Sound quality varies dramatically depending on the voice characteristics of the wearer, where the device is worn, and how it is attached. The device must be properly secured with adequate pressure and the amplifier must be dynamically configured to achieve optimal signal quality.

b) Reliability: The physiological microphone must remain firmly in contact with the wearer's skin in order to transmit speech. No speech is transmitted if contact is interrupted or lost.

c) Availability and Cost: An amplifier is required to provide an interface between the microphone and the radio. The amplifier requires custom configuration to connect to conventional radios and is not commercially available. Further research is required to evaluate the physiological microphone's effectiveness for speech communications.

2.6 Lip Tracking Continuous Speech Recognition

Speech recognition has been gaining wide acceptance as a computer input modality in recent years. The availability of a number of commercial, off-the-shelf ASR (Active Speech Recognition) products is the result of decades of research plus the fact that personal computers are becoming faster and more affordable. However, it remains a challenge for ASR systems to perform robustly in adverse environments such aircrafts' cockpits, battlefields, or other scenarios where background noise would be an issue.

One approach being exploring is the use of visual cues from the speaker lips to augment the capability of acoustic-based speech recognition systems. An advantage of bimodal approach is that information from lip movements are often complementary to that of the acoustic counterpart, as was confirmed by previous psychophysical studies. Secondly, accuracy of visual feature measurements is unaffected by acoustic noise, regardless of the level of signal-to-noise ratio. Supported in part by the Army Research Laboratory, Rockwell Scientific developed an integrated system designed to automatically localize the speaker's lips and track their movements at video rate. Researchers are currently developing a real-time bimodal recognition system.

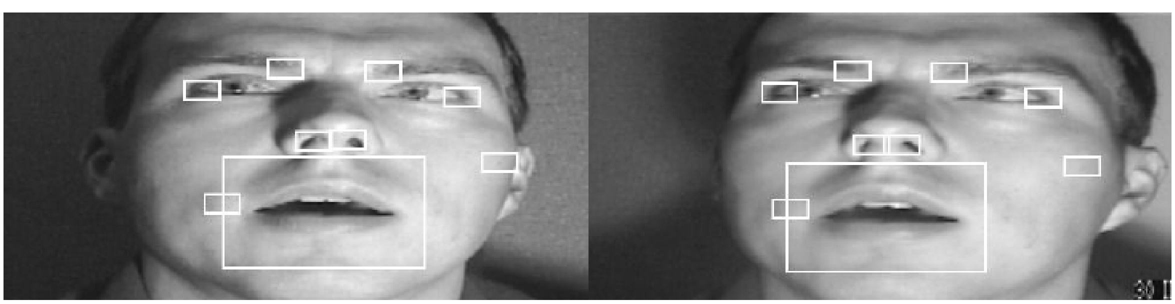

Figure 6. Stereo Image with Templates and Mouth Windows

2.6.1 Advantages

a) Free Field Sound Detection: The camera boom does not interfere with the wearer's ability to detect sounds in the free field.

b) Noise Attenuation: Lip tracking is completely resistant to background noise in the free field.

c) Fit/Comfort: The lip tracker is positioned away from the wearer's face. The camera boom can be moved away from the wearer's face when it is not in use. Lip tracking cameras are small and light enough to be worn for an extended period of time.

d) Tactical Feasibility: The lip-tracking camera can detect silent or whispered speech for tactical use.

2.6.2 Disadvantages

a) Speech Input Performance: Lip tracking alone is not adequate to detect speech. Lip tracking is intended for use in conjunction with an acoustic or transmissive speech input device. Digital pattern recognition is used to detect speech. Speech recognition accuracy is measurably improved when an acoustic or transmissive speech input device is used in conjunction with lip tracking. In a speaker-dependent experiment by Rockwell Scientific involving a moderate-size vocabulary, the hidden Markov model-based bimodal speech recognizer was shown to outperform its acoustic-only counterpart, with error reduction rate of up to 30%. With a small, restricted vocabulary, the system developed by Rockwell was able to perform recognition using information from lip motion alone. By integrating geometric features with properly normalized pixel-based features, recognition accuracy up to 98% was achievable in an isolated digit recognition task.

b) Equipment Compatibility: The lip-tracking camera must be attached to the wearer's head and the mouth must be visually unobstructed. This device is incompatible with balaclavas, face masks, gas masks, and respiration devices. The camera boom can interfere with the soldier's helmet or rifle.

c) Fit/Comfort: The camera boom can brush against the wearer's face.

d) Reliability: Soldiers must ensure the camera is positioned in front of their mouth to track the lip position. No speech will be captured if the camera is not aligned to the wearer's lips. The camera requires light in order to detect the position of the wearer's lips. Shadows or reduced light conditions may interfere with speech detection accuracy.

e) Durability: The camera is a fragile device that can be damaged by impact or shock. Dust or debris in the environment can obstruct the camera lens.

f) Availability and Cost: Wearable lip tracking speech input devices are not commercially available at any price. Rockwell developed these devices for experimentation purposes.

3. Results and Discussion

Weights were systematically assigned to each of the performance rating criteria. A perfect score (i.e. 7 points out of a possible 7) was allocated to devices that provided a distinct or unique advantage and no disadvantages for a given category. Deductions were made to scores based on the perceived severity of the apparent disadvantages identified. Performance ratings assigned to each device were tallied and devices were ranked according to their total scores.

Results indicate certain speech input devices are more compatible with soldiering tasks. Table 2 provides a ratings summary for each of the input device alternatives according to the performance criteria used to evaluate each of the systems. Overall, the bone conduction microphone is rated better than other types of speech input devices. While the bone conduction microphone has no unique advantages over other devices in terms of Speech Input Performance, Equipment Compatibility, Fit and Comfort, Tactical Feasibility, Reliability, and Availability and Cost, it achieves consistently high ratings in terms of Free Field Sound Detection, Noise Attenuation, and Durability.

The physiological microphone was rated second most acceptable overall. It has unique advantages over all other headsets in terms of Equipment Compatibility and Fit and Comfort. If a reliable COTS (Commercial Off-The-Shelf) version of the physiological microphone were available with superior audio quality, it would likely achieve the highest overall rating.

The boom microphone was rated third most acceptable overall. The boom microphone also has unique advantages in terms of Speech Input Performance, Reliability, and Availability and Cost. The boom is the most familiar and conventional type of headset. It is durable and accurate but is susceptible to free field sounds.

The bone conduction, physiological, and lip tracking input devices were all rated best in terms of Free Field Noise Attenuation. The bone conduction and physiological microphones are rated better than the other display types in terms of Durability. The in-ear microphone was the only headset that did not receive the highest possible acceptability rating for Free Field Sound Detection.

Table 3. Performance Measures for Speech Input Devices

Rating Criteria	Boom Microphone	Throat Microphone	In-Ear Microphone	Bone Conduction Microphone	Physiological Microphone	Lip Tracking Continuous Speech Recognition
Free Field Sound Detection	7	7	3	7	7	7
Speech Input Performance	6	4	5	5	1	2
Noise Attenuation	3	6	6	7	7	7
Equipment Compatibility	4	6	6	6	7	2
Fit & Comfort	4	4	4	6.5	7	4
Tactical Feasibility	5	5	5	6	5.5	7
Reliability	6	3	4	4	4	2
Durability	5	2	2	6	6	2
Availability & Cost	7	6	4	4	3	1
Total	47	43	39	51.5	47.5	34

4. References

BOS, J.C. and TACK, D.W., 2005. Input Device Investigation for Future Dismounted Soldier Computer Systems. *CR2005-052*. Toronto:ON, *Defence Research and Development Canada – Toronto*.

COLBERT, H , TACK, D.W., Bos, J.C., 2005. Radio Communication Input and Display Devices for Infantry Soldiers. *CR2005-039.* Toronto:ON, *Defence Research and Development Canada –Toronto*.

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document, Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's document, or tasking agency, are entered in section 8.)

Publishing: DRDC Toronto

Performing: Humansystems® Incorporated, 111 Farquhar St., 2nd floor, Guelph, ON N1H 3N4

Monitoring:

Contracting: DRDC Toronto

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parenthesis at the end of the title)

Speech Input Hardware Investigation for Future Dismounted Soldier Computer Systems (U)

4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.)

Jeffrey C. Bos ; David W. Tack

5. DATE OF PUBLICATION (Month and year of publication of document.)

May 2005

6a NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)

26

6b. NO. OF REFS (Total cited in document.)

2

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of document, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development − include address.)

Sponsoring: DLR 5, NDHQ OTTAWA,ON K1A 0K2

Tasking:

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.)

12QG01

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

W7711−017747/001/TOR

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document)

DRDC Toronto CR 2005−064

10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.)

SIREQ #152

11. DOCUMENT AVAILABILITY (Any limitations on the dissemination of the document, other than those imposed by security classification.)

Defence departments in approved countries Document has initial limited distribution through Exploitation Manager TTCP and NATO countries and agencies Unlimited after initial limited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11), However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Other Document to have initial limited announcement

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) Speech input devices are required by future infantry solders to provide voice input to radio communications systems for voice communication between individuals and to provide speech input to electronic speech recognition systems. A comparison of six types of audio input devices (bone conduction, in−ear, boom, throat and physiological microphones and lip tracking continuous speech recognition) was conducted based upon existing literature, user feedback and the authors' experience. This report presents the results of that comparison. The results indicate some devices are more compatible with certain soldiering tasks. Overall, the bone conduction microphone is rated better than other types of speech input devices. While the bone conduction microphone has no unique advantages over the other devices in terms of Speech Input Performance, Equipment Compatibility, Fit and Comfort, Tactical Feasibility, Reliability, and Availability and Cost, it achieves consistently high ratings in terms of Free Field Sound Detection, Noise Attenuation, and Durability. The physiological microphone was rated second most acceptable overall. The physiological microphone has unique advantages over all other headsets in terms of Equipment Compatibility and Fit and Comfort. The physiological microphone would likely achieve the highest overall rating if a reliable COTS (Commercial Off−The−Shelf) version was available with acceptable audio quality. The boom microphone was rated third most acceptable overall. The boom microphone also has unique advantages in terms of Speech Input Performance, Reliability, and Availability and Cost. The boom microphone is the most familiar and conventional headset. It is durable and accurate but is very susceptible to free field noise.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) Soldier Information Requirements Technology Demonstration Project; SIREQ TD; speech input device; radio communications; electronic speech recognition system; bone conduction microphone; physiological microphone; boom microphone