6. Promote the students' willingness to speak, by making the student feel that the teacher is interested in *what* the student has to say and not only by *how* it is said.
7. Provide positive feedback when the student has made an effort or when a progress is made.
8. Adapt to the exercise. Use explicit feedback sparingly if implicit feedback is enough.
9. Give feedback only on the focus of the session. If other pronunciation problems are discovered, these should be left uncorrected, but noted and addressed in another session.

5 Feedback management in ARTUR

Some aspects of the feedback strategies proposed above have been implemented in a Wizard-of-Oz version of ARTUR that will be demonstrated at the conference. The focus of the exercise is to teach speakers of English the pronunciation of the Swedish sound "sj", using the tongue twister "Sju själviska sjuksköterskor stjäl schyst champagne".

The instructions and feedback consisted of instructions and animations on how to position the tongue, showing and explaining the difference between the user's pronunciation and the correct. The user could further listen to his/her previous attempt to compare it with the target.

One new feature is that each user can control individually the amount of feedback given. The first reason for this is the affective, that students should be able to choose a level that they are comfortable with. The second is that this does put the responsibility and initiative with the student, who can decide how much advice he or she requires from the tutor.

Secondly, several feedback categories have been added to the standard *positive* (for a correct pronunciation) and *corrective* (incorrect): *minimal* (correct pronunciation, only implicit positive feedback given, in order not to interrupt the flow of the training), *satisfactory* (the pronunciation is not entirely correct, but it is pedagogically sounder to accept it and move ahead), *augmented* (for a repeated error, more detailed feedback given), *vague* (a general hint is given, rather than explicit feedback) and *encouragement* (encouraging the student and asking for a new try). The two latter categories may be used either when the system is uncertain of the error, when it does not fit the predefined mispronunciation categories or when more explicit feedback is pedagogically unsound.

Acknowledgements

This research is carried out within the ARTUR project, funded by the Swedish research council. The Centre for Speech Technology is supported by VINNOVA (The Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. The author would like to thank the participating teachers and students.

References

Bälter, O., O. Engwall, A-M. Öster & H. Kjellström, 2005. Wizard-of-oz test of ARTUR – a computerbased speech training system with articulation correction. *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 36–43.

Carroll, S. & M. Swain, 1993. Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Lang. Acquisition 15*, 357–386.

Lyster, R. & L. Ranta, 1997. Corrective feedback and learner uptake. *Studies in Second Lang. Acquisition 20*, 37–66.

Mackey, A. & J. Philip, 1998. Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal 82*, 338–356.

Neri, A., C. Cucchiarini & H. Strik, 2002. Feedback in computer assisted pronunciation training: When technology meets pedagogy. *Proceedings of CALL professionals and the future of CALL research*, 179–188.

Rubin, J. (ed.), 1994. *Handbook of Usability Testing*. New York: John Wiley & Sons Inc.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 45–48

Directional Hearing in a Humanoid Robot

Evaluation of Microphones Regarding HRTF and Azimuthal Dependence

Lisa Gustavsson, Ellen Marklund, Eeva Klintfors, and Francisco Lacerda
Department of Linguistics/Phonetics, Stockholm University
{lisag|ellen|eevak|frasse}@ling.su.se

Abstract

As a first step of implementing directional hearing in a humanoid robot two types of microphones were evaluated regarding HRTF (head related transfer function) and azimuthal dependence. The sound level difference between a signal from the right ear and the left ear is one of the cues humans use to localize a sound source. In the same way this process could be applied in robotics where the sound level difference between a signal from the right microphone and the left microphone is calculated for orienting towards a sound source. The microphones were attached as ears on the robot-head and tested regarding frequency response with logarithmic sweep-tones at azimuth angles in 45° increments around the head. The directional type of microphone was more sensitive to azimuth and head shadow and probably more suitable for directional hearing in the robot.

1 Introduction

As part of the CONTACT project[1] a microphone evaluation regarding head related transfer function (HRTF), and azimuthal[2] dependence was carried out as a first step in implementing directional hearing in a humanoid robot (see Figure 1). Sound pressure level by the robot ears (microphones) as a function of frequency and azimuth in the horizontal plane was studied.

The hearing system in humans has many features that together enable fairly good spatial perception of sound, such as timing differences between left and right ear in the arrival of a signal (interaural time difference), the cavities of the pinnae that enhance certain frequencies depending on direction and the neural processing of these two perceived signals (Pickles, 1988). The shape of the outer ears is indeed of great importance in localization of a sound source, but as a first step of implementing directional hearing in a robot, we want to start up by investigating the effect of a spherical head shape between the two microphones and the angle in relation to the sound source. So this study was done with reference to the interaural level difference (ILD)[3] between two ears (microphones, no outer ears) in the sound signal that is caused by the distance between the ears and HRTF or head shadowing effects (Gelfand, 1998). This means that the ear furthest away from the sound source will to some extent be blocked by the head in such a way that the shorter wavelengths (higher frequencies) are reflected by the head (Feddersen et al., 1957). Such frequency-dependent differences in intensity associated with different sound source locations will be used as an indication to the robot to turn his head in the horizontal plane. The principle here is to make the robot look in the direction that minimizes the ILD[4]. Two types of microphones, mounted on the robot head,

[1] "Learning and development of Contextual Action" European Union NEST project 5010
[2] Azimuth = angles around the head
[3] The abbreviation IID can also be found in the literature and stands for Interaural Intensity Difference.
[4] This is done using a perturbation technique. The robot's head orientation is incrementally changed in order to detect the direction associated with a minimum of ILD.

were tested regarding frequency response at azimuth angles in 45° increments from the sound source (Shaw & Vaillancourt, 1985; Shaw, 1974).

The study reported in this paper was carried out by the CONTACT vision group (Computer Vision and Robotics Lab, IST Lisbon) and the CONTACT speech group (Phonetics Lab, Stockholm University) assisted by Hassan Djamshidpey and Peter Branderud. The tests were performed in the anechoic chamber at Stockholm University in December 2005.

2 Method

The microphones evaluated in this study were wired Lavalier microphones of the Microflex MX100 model by Shure. These microphones were chosen because they are small electret condenser microphones designed for speech and vocal pickup. The two types tested were omni-directional (360°) and directional (cardoid 130°). The frequency response is 50 to 17000 Hz and its max SPL is 116 dB (omni-directional), 124 dB (directional) with a s/n ratio of 73 dB (omni-directional), 66 dB (directional). The robotic head was developed at Computer Vision and Robotics Lab, IST Lisbon (Beira et al., 2006).

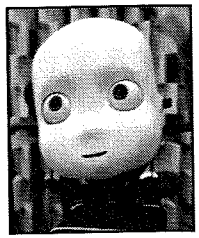

Figure 1. Robot head.

2.1 Setup

The experimental setup is illustrated in Figure 2a. The robot-head is attached to a couple of ball bearing arms (imagined to correspond to a neck) on a box (containing the motor for driving head and neck movements). The microphones were attached and tilted by about 30 degrees towards the front, with play-dough in the holes made in the skull for the future ears of the robot. The wires run through the head and out to the external amplifier. The sound source was a Brüel&Kjær 4215, Artificial Voice Loud Speaker, located 90 cm away from the head in the horizontal plane (Figures 2a and 2b). A reference microphone was connected to the loudspeaker for audio level compression (300 dB/sec).

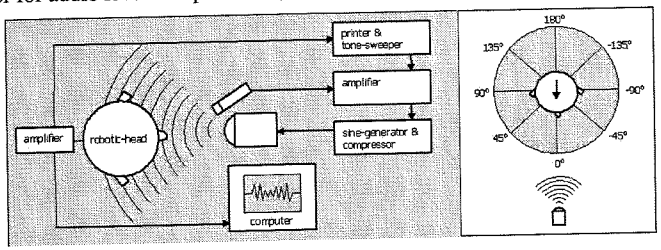

Figure 2a and 2b. a) Wiring diagram of experimental setup (left). b) Azimuth angles in relation to robot head and loudspeaker (right).

2.2 Test

Sweep-tones were presented through the loud-speaker at azimuth angles in 45° increments obtained by turning the robot-head (Figure 2b). The frequency range of the tone was between 20 Hz[5] and 20 kHz with a logarithmic sweep control and writing speed of 160mm/sec (approximate averaging time 0.03 sec). The signal response of the microphones was registered and printed in dB/Hz diagrams (using Brüel&Kjær 2307, Printer/Level Recorder) and a back-up recording was made on a hard-drive. The dB values as a function of frequency were also plotted in Excel diagrams for a better overview of superimposed curves of different azimuths (and for presentation in this paper).

[5] Because the compression was unstable up until about 200 Hz, the data below 200 Hz will not be reported here. Furthermore, the lower frequencies are not affected that much in terms of ILD.

3 Results

The best overall frequency response of both microphones was at angles 0°, -45° and -90° that is the (right) microphone is to some extent directed towards the sound source. The sound level decreases as the microphone is turned away from the direction of the sound source. The omni-directional microphones have an overall more robust frequency response than the directional microphones. As expected, the difference in sound level between the azimuth angles are most significant in higher frequencies since the head as a blockage will have a larger impact on shorter wavelengths than on longer wavelengths. An example of ILD for the directional microphones is shown in Figure 3, where sound level as a function of frequency is plotted for the ear near the sound source and ear furthest away from the sound source at azimuth 45°, 90° and 135°. While the difference ideally should be zero[6] at azimuth 0 it is well above 15 dB at many higher frequencies in azimuth 45°, 90° and 135°.

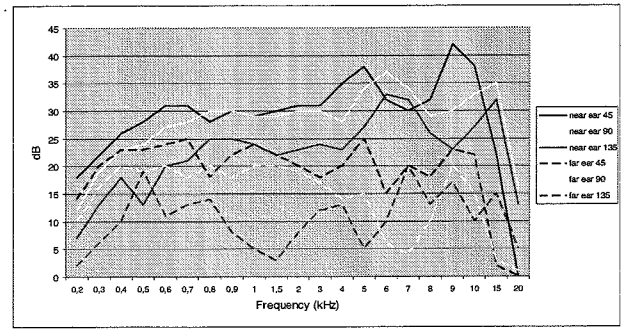

Figure 3. Signal response of the directional microphones. Sound level as a function of frequency and azimuth 45°, 90° and 135° for the ear near the sound source and the ear furthest away from the sound source.

4 Discussion

In line with the technical description of the microphones our results show that the directional microphones are more sensitive to azimuth than the omnidirectional microphones and will probably make the implementation of sound source localization easier. Also disturbing sound of motors and fans inside the robot's head might be picked up easier by an omnidirectional microphone. A directional type of microphone would therefore be our choice of ears for the robot. However, decisions like this are not made without some hesitation since we do not want to manipulate the signal response in the robot hearing mechanism beyond what we find motivated in terms of the human physiology of hearing. Deciding upon what kind of pickup angle the microphones should have forces us to consider what implications a narrow versus a wide pickup angle will have in further implementations of the robotic hearing. At this moment we see no problems with a narrow angle but if problems arise we can of course switch to wide angle cartridges.

The reasoning in this study holds for locating a sound source only to a certain extent. By calculating the ILD the robot will be able to orient towards a sound source in the frontal horizontal plane. But if the sound source is located straight behind the robot the ILD would also equal zero and according to the robot's calculations he is then facing the sound source. Such front-back errors are in fact seen also in humans since there are no physiological

[6] A zero difference in sound level at all frequencies between the two ears requires that the physical surroundings at both sides of the head are absolutely equal.

attributes of the ear that in a straightforward manner differentiate signals from the front and rear. Many animals have the ability to localize a sound source by wiggling their ears, humans can instead move themselves or the head to explore the sound source direction (Wightman & Kistler, 1999). As mentioned earlier the outer ear is however of great importance for locating a sound source, the shape of the pinnae does enhance sound from the front in certain ways but it takes practice to make use of such cues. In the same way the shape of the pinnae can be of importance for locating sound sources in the medial plane (Gardner & Gardner, 1973; Musicant & Butler, 1984). Subtle movements of the head, experience of sound reflections in different acoustic settings and learning how to use pinnae related cues are some solutions to the front-back-up-down ambiguity that could be adopted also by the robot. We should not forget though, that humans always use multiple sources of information for on-line problem solving and this is most probably the case also when locating sound sources. When we hear a sound there is usually an event or an object that caused that sound, a sound source that we easily could spot with our eyes. So the next question we need to ask is how important vision is in localizing sound sources or in the process of learning how to trace sound sources with our ears and how vision can be used in the implementation of directional hearing of the robot.

5 Concluding remarks

Directional hearing is only one of the many aspects of human information processing that we have to consider when mimicking human behaviour in an embodied robot system. In this paper we have discussed how the head has an impact on the intensity of signals at different frequencies and how this principle can be used also for sound source localization in robotics. The signal responses of two types of microphones were tested regarding HRTF at different azimuths as a first step of implementing directional hearing in a humanoid robot. The next steps are designing outer ears and formalizing the processes of directional hearing for implementation and on-line evaluations (Hörnstein et al., 2006).

References

Beira, R., M. Lopes, C. Miguel, J. Santos-Victor, A. Bernardino, G. Metta et al., in press. Design of the robot-cub (icub) head. *IEEE ICRA*.

Feddersen, W.E., T.T. Sandel, D.C. Teas & L.A. Jeffress, 1957. Localization of High-Frequency Tones. *Journal of the Acoustical Society of America 29*, 988-991.

Gardner, M.B. & R.S. Gardner, 1973. Problem of localization in the median plane: effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America 53*, 400-408.

Gelfand, S., 1998. *An introduction to psychological and physiological acoustics*. New York: Marcel Dekker, Inc.

Hörnstein, J., M. Lopes & J. Santos-Victor, 2006. Sound localization for humanoid robots – building audio-motor maps based on the HRTF. *CONTACT project report*.

Musicant, A.D. & R.A. Butler, 1984. The influence of pinnae-based spectral cues on sound localization. *Journal of the Acoustical Society of America 75*, 1195-1200.

Pickles, J.O., 1988. *An Introduction to the Physiology of Hearing*. (Second ed.) London: Academic Press.

Shaw, E.A.G., 1974. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *Journal of the Acoustical Society of America 56*.

Shaw, E.A.G. & M.M. Vaillancourt, 1985. Transformation of sound-pressure level from the free field to the eardrum presented in numerical form. *Journal of the Acoustical Society of America 78*, 1120-1123.

Wightman, F.L. & D.J. Kistler, 1999. Resolution of front—back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America 105*, 2841-2853.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 49–52

Microphones and Measurements

Gert Foget Hansen[1] and Nicolai Pharao[2]
[1]Department of Dialectology, University of Copenhagen
gertfh@hum.ku.dk
[2]Centre for Language Change in Real Time, University of Copenhagen
nicolajp@hum.ku.dk

Abstract

This paper presents the current status of an ongoing investigation of differences in formant estimates of vowels that may come about solely due to the circumstances of the recording of the speech material. The impact of the interplay between type and placement of microphone and room acoustics are to be examined for adult males and females across a number of vowel qualities. Furthermore, two estimation methods will be compared (LPC vs. manual). We present the pilot experiment that initiated the project along with a brief discussion of some relevant articles. The pilot experiment as well as the available results from other related experiments seem to indicate that different recording circumstances could induce apparent formant differences of a magnitude comparable to differences reported in some investigations of sound change.

1 Introduction

1.1 Purpose

The study reported here arose from a request to evaluate different types of recording equipment for the LANCHART Project, a longitudinal study of language change with Danish as an example. One aim of the assignment was to ensure that the LANCHART corpus would be suitable for certain acoustic phonetic investigations.

1.2 Pilot experiments – choosing suitable microphones for on-location recordings

Head mounted microphones were compared to the performance of a lapel-worn microphone and a full-size directional microphone placed in a microphone stand in front of the speaker (hereafter referred to as a studio microphone). The following four factors were considered in the evaluation of the suitability of the recordings provided by the microphones: 1) ease of transcription and 2) segmentation of the recordings as well as estimation of 3) fundamental frequency and 4) formants using LPC analysis.

Simultaneous recordings of one speaker using all three types of microphones formed the basis for a pilot investigation. Primarily, the results indicated that the lapel-worn microphone was clearly inferior to the other two types with regard to the first 3 criteria, since it is more prone to pick up background noise. The head mounted and studio microphones also showed some differences with regard to these 3 criteria; in particular the recordings made with the head mounted microphone provided clearer spectrograms. Furthermore, apparent differences emerged in the LPC analysis of the vowels in the three recordings.

To explore this further we recorded 6 different pairs of microphone and distance combinations using a two channel hard disk recorder. Microphones compared were Sennheiser ME64, Sennheiser MKE2 lavallier and VT600 headset microphone, positioned either as indicated by type, or as typical for ME64 (i.e. at a distance of about 30 cm).