(loosely based on Mermelstein, 1975) contained in it. A hull in the intensity values of speech
is assumed to correspond roughly to a syllable, thus providing a pseudo-syllabification, or
syllablification. By searching backwards, the hull that occurred last is found first. Currently,
processing ceases at this point, since only the hulls directly preceding silence has been of
interest to us so far. A convex hull in /naɪlɒn/ is defined as a stretch of consecutive value
triplets ordered chronologically, where the centre value is always above or on a line drawn
between the first and the last value. As this definition is very sensitive to noisy data, it is
relaxed by allowing a limited number of values to drop below the line between first and last
value as long as the area between that line and the actual values is less than a preset threshold.

3.8 Classification
The normalised pitch, intensity, and voicing data extracted by /naɪlɒn/ over a psyllable are
intended for classification of intonation patterns. Each silence-preceding hull is classified into
HIGH, MID, or LOW depending on whether the pitch value is in the upper, mid or lower third of
the speaker's F0 range described by mean and standard deviation, and into RISE, FALL, and or
LEVEL depending on the shape of the intonation pattern. Previous work has shown that the
prosodic information provided by /naɪlɒn/ can be used to improve the interaction control in
spoken human-computer dialogue compared to systems relying exclusively on silence
duration thresholds (Edlund & Heldner, 2005).

4 Discussion
In this paper, we have presented /naɪlɒn/, an online, real-time software package for prosodic
analysis capturing a number of prosodic features liable to be relevant for interaction control.
Future work will include further development of /naɪlɒn/ in terms of improving existing
algorithms — in particular the intonation pattern classification — as well as adding new
prosodic features. For example, we are considering evaluating the duration of psyllables as an
estimate of final lengthening or speaking rate effects, and to use intensity measures to capture
the different qualities of silent pauses resulting from different vocal tract configurations
(Local & Kelly, 1986).

Acknowledgements
This work was carried out within the CHIL project. CHIL is an Integrated Project under the
European Commission's sixth Framework Program (IP-506909).

References
226.
Ferrer, L., E. Shrøberg & A. Stolcke, 2002. Is the speaker done yet? Faster and more accurate
end-of-utterance detection using prosody in human-computer dialog. *Proceedings ICISLP*
2002, Denver, 2061-2064.
structure. *Human Studies* 9, 185-204.
Mermelstein, P., 1975. Automatic segmentation of speech into syllabic units. *Journal of the
Acoustical Society of America* 58, 880-883.
Ward, N. & W. Tsukahara, 2000. Prosodic features which cue back-channel responses in
2 Feedback in pronunciation training

Lyster & Ranta (1997) classified feedback given by language teachers as

1. **Explicit correction**: The teacher clearly states that what the student said was incorrect and gives the correct form, e.g. as “You should say: ...”

2. **Recasts**: The teacher reformulates the student's utterance, removing the error.

3. **Repetition**: The teacher repeats the student utterance with the error using the intonation to indicate the error. Repetitions may also be used as positive feedback on a correct utterance.

4. **Clarification requests**: Urging the student to reformulate the utterance.

5. **Metalinguistic feedback**: Information or questions about an error used to make the students reflect upon and find the error themselves using the provided information.

6. **Elicitation**: Encourage students to provide the correct pronunciation, by open-ended questions or fill-in-the-gap utterances.

Recasts were by far the most common type, but learners often perceive recasts as another way to say the same thing, rather than a correction (Mackey & Philip, 1998). Carroll & Swain (1993) found that all groups receiving feedback, explicit or implicit, improved significantly more than the control group, but the group given explicit feedback outperformed the others. As explicit feedback may be intrusive and affect student self-confidence if given too frequently, it is however not evident that it should always be used.

3 Data collection

Six language teachers participated in the study, four in a focus group and two in individual interviews using a semi-structured protocol (Rubin, 1994) with open-ended questions. Five students were interviewed, three of them in a focus group and two individually. The teacher and student groups were intentionally heterogeneous with respect to target language and student level, in order to capture general pedagogical strategies. Classroom observations were made in three beginner level courses, where the languages taught were close to, moderately and student groups were intentionally heterogeneous with respect to target language and student level, in order to capture general pedagogical strategies. Classroom observations were made in three beginner level courses, where the languages taught were close to, moderately different from and very different from Swedish, respectively.

4 Results

4.1 **When should errors be corrected?**

There was a large consensus among teachers and students about the importance of never interrupting the students' utterances, reading or discussions with feedback, even if it means that errors are left uncorrected. This strategy was also observed in the classrooms.

4.2 **How should errors be corrected?**

This section summarizes how the teachers (T) or students (S) described how feedback should be given and feedback observed during classes (O).

1. **Recasts** were the most common feedback in the classroom and were also advocated by the students, as they considered that it was often enough to hear the correct pronunciation. Contrary to the finding by Mackey & Philip (1998) that recasts were not perceived as corrections, the students tried to repair after recasts (T, S, O).

2. **Implicit** (e.g. “Sorry?”) and explicit (e.g. “Could you repeat that?”) elicitation for the student to self-correct was used frequently (O).

3. **Increasing feedback**. One teacher described a strategy going from minimal implicit feedback towards more explicit, when required. In the most minimal form, the teacher indicates that an error was produced by a questioning look or an attention-catching sound, giving the students the opportunity to identify and self-correct the error. If the student is unable to repair, a recast would be used. If needed, the recast would be repeated again (turning it into an explicit correction). The last step would be an explicit explanation of the difference between the correct and erroneous pronunciation (T).

4. **Articulatory instructions**. Several teachers thought that formal descriptions and sketches on place of articulation are of little use, since the students are unaccustomed to thinking about how to produce different sounds. Some teachers did, however, use articulatory instructions and one student specifically requested this type of feedback (T, S, O).

5. **Sensory feedback**, e.g. letting the students place their hands on their neck to feel the vibration of voiced sounds or in front of the mouth to feel aspiration (T, O).

6. **Comparisons to Swedish phonemes**, as an approximation or reminder (T, S, O).

7. **Metalinguistic explanations** used to enforce the feedback or to motivate why it is important to get a particular pronunciation right (T).

8. **General recommendations** rather than feedback on particular errors, e.g., “You should try reading aloud by yourself at home”, to encourage additional training (T, O).

9. **Contrasting repeat-recast**, to illustrate the difference between the student utterance and the correct or between minimal pairs (T, S).

4.3 **Which errors should be corrected?**

The teachers ventured several criteria for which errors should be corrected:

1. **Comprehensibility**: If the utterance could not be correctly understood.

2. **Intelligibility**: If the utterance could not be understood without effort.

3. **Frequency**: If the student repeats the same (type of) error several times.

4. **Social impact**: If the listener gets a negative impression of a speaker making the error.

5. **Proficiency**: If a student with a better overall pronunciation may get corrective feedback on an error for which a student with a less good pronunciation does not get one.

6. **Generality**: If the error is one that is often made in the L2 by foreign speakers.

7. **Personalisation**: A student who appreciates corrections receives more than one who does not.

8. **Commonality**: An error that is common among native speakers of the L2 language is regarded as less grave than such errors that a native speaker would never make.

9. **Exercise focus**: Feedback is primarily given on the feature targeted by the exercise.

None of the students thought that all errors should be corrected, only the “worst”. When questioned further, the general opinion was that this signified mispronunciations that lead to misunderstandings or deteriorating communication. Other criteria stated were if the error affected the listener’s view of the speaker negatively, or if it was a repeated error. Apart from this, the students thought that it should depend on the student’s ambition. These opinions hence correspond to the first five criteria given by the teachers.

In the classes, the amount and type of feedback given depended on the type of exercise (practicing one word, reading texts, speaking freely), the L2 language (for the L2 language that was most different from Swedish, significantly more detailed feedback was given), generality (errors that several students made were given more emphasis) and proficiency.

4.4 **Motivation**

To avoid negative feelings about feedback, the teachers or students suggested:

1. Adapt the feedback to the students' self-confidence (criteria 5 & 7 in section 3.3).

2. Make explicit corrections impersonal, by expanding to a general error and using “When one says...” rather than “When you say...”

3. Insert non-problematic pronunciations among the more difficult ones.

4. Acknowledge difficulties (e.g. “Yes, this is a tricky pronunciation”).

5. Never getting stuck on the same pronunciation too long.
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 sjalviska sjukaköterskor stjal 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) 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

Lisa Gustavsson, Ellen Marklund, Eva Klintfors, and Francisco Lacerda
Department of Linguistics/Phonetics, Stockholm University
{lisa@ellen@evas@frassa}@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 project1 a microphone evaluation regarding head related transfer function (HRTF), and azimuthal2 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 (microphones) as a function of frequency and azimuth in the horizontal plane was studied.

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 ILD4. Two types of microphones, mounted on the robot head,