How to tell beans from farmers: cues to the perception of pitch accent in whispered Norwegian

Hannele Nicholson and Andreas Hilmo Teig
Phonetics Laboratory, University of Oxford

1. Introduction

Though quite distinct in physical form, 'beans' and 'farmers' are distinguishable lexically only by a difference in pitch accent in spoken Norwegian. In the absence of pitch information, Fintoft (1970) investigates whether or not listeners are able to discern between the two tokens in whispered speech – without the assistance of context. Though his findings suggest that listeners may rely upon the presence of an additional cue present in the whispered speech stream, the possibility that context could aid the listeners in detecting the appropriate pitch accent is not discussed. This paper presents the results of an experiment conducted to assess listeners' ability in determining which pitch accent word token best fits into a whispered ambiguous utterance in spoken Norwegian. The results confirm that context is not a reliable cue to assist in lexical selection and concur with Fintoft (1970) in suggesting that listeners utilise a separate prosodic cue, possibly syllable duration or intensity, to make the pitch accent distinction in whispered speech.

2. The Problem

East Norwegian employs pitch accent contours in order to make lexical distinctions. For example, the words /[1]bøner/ **'***farmers'* and /[2]bøner/ '*beans'*[1] consist of identical phonetic segments. The only way to distinguish between these words are to rely on a) the context within which the utterance is created and/or b) the pitch accent accompanying the word. This paper seeks to investigate whether there is enough information present in the speech signal for the speakers not to have to rely on any of these, say during whisper.

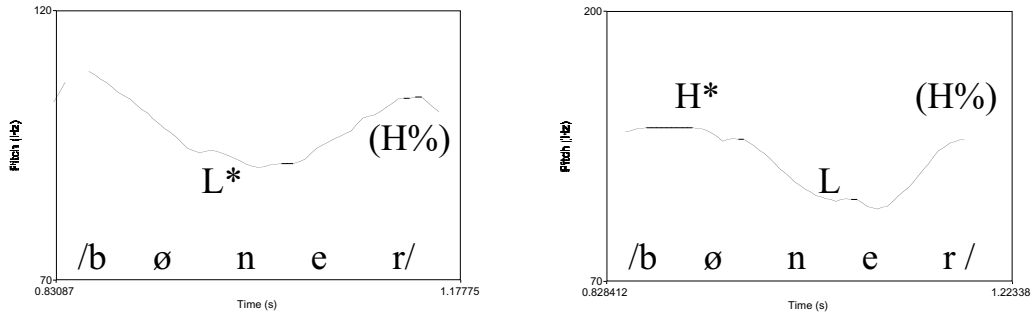
From Figure 1, we can see that pitch accent 1 is associated with a simple lexical L*, here followed by a H% boundary tone[2]. Pitch accent two, on the other hand, consists of a lexical H* and a L, here occurring with a H% boundary tone. There are some 3000 pairs of pitch accent words

[1] Accents 1 and 2 are normally marked with a superscript 1 and 2 in Norwegian.

[2] This follows the system of Norwegian intonation called the Trondheim model, described in e.g. Nilsen 1992.

Nordlyd 31.2:315-325, © the author 2003
Proceedings of the 19th Scandinavian Conference of Linguistics, vol. 31.2,
Edited by Anne Dahl, Kristine Bentzen, and Peter Svenonius

in the language, each of which are disambiguated in this way (Fintoft 1970).



'farmers' *'beans'*

Figure 1. *Pitch traces with typical intonation contours for two words distinguished only by accent 1 (left) and accent 2 (right).*

The second part of the phenomenon is explainable in phonetic terms. During whispered phonation, the vocal folds do not vibrate (see e.g. Catford 1977). This is due to properties of the glottis, whereby in order for whispered phonation to be produced, the glottis is considerably narrowed at the anterior end. This narrowing allows a turbulent air stream to pass through the glottis without the generation of voicing. Figure 2, from Saunders (2002), portrays a comparative picture of the glottis during voiced and whispered speech.

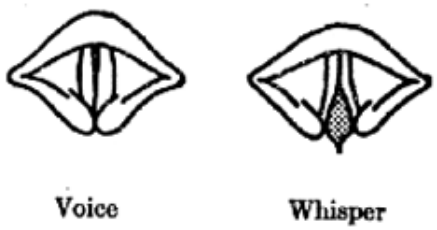


Figure 2. *A view of a the glottis during voiced speech and whispered speech.*

During pitch production, it has long been known that the vocal folds must vibrate in order to produce pitch (Lass 1996). The vocal folds open and close at a periodic rate, known as fundamental frequency, or f_0 (see e.g. Ladefoged 2001). Therefore, it is physically impossible for a speaker to produce whispered speech if the vocal folds are vibrating. If, on the other hand, f_0 is not present, there is no pitch and the listener must rely on either the linguistic or extra-linguistic context or some other acoustic cue to distinguish between pitch accent contours in Norwegian. In his seminal

study of Norwegian pitch accents, Fintoft (1970) investigated the issue for whispered Norwegian. However, his study concerned only identification of pitch accent in single word tokens spoken out of context.

2.1 Is whisper perceptible?

The fact that pitch information is dependent upon f_0 information leads one to wonder how languages that consistently employ tone make lexical distinctions function in the absence of pitch information, such as in whispered speech. This question has been researched before for Chinese, Thai, Norwegian and Swedish (Giet 1956). No matter what language is used as a means of investigating the issue, the conclusion is the same. Pitch is perceptible in whispered speech. There is much debate however, over the mechanism, whether contextual or prosodic , that may act as a substitute for voicing and pitch.

Meyer-Eppler (1957) suggests that intonational information is understood via a spectral shift in whispered speech. To research this matter, Meyer-Eppler (1957) conducted an acoustic analysis of German vowels that had been "sung" without voicing. For front vowels, a second formant frequency (F2) was detectable at 2000 Hz. The first formant (F1) of back vowels exhibited a value much higher than that of a normal voiced vowel. The validity of Meyer-Eppler's results is somewhat contestable, in that requesting subjects to "sing" without voicing is a rather unnatural, non-speech like act.

A second observation, also made on the basis of evidence from whispered vowels, suggest that listeners perceive a pitch in the F2 range (Thomas 1969). Three musically trained listeners were presented with a variety of whispered stimuli and asked to adjust an oscillator dial to within 10 Hz of the perceived pitch. This task was repeated twice for each of the three listeners and listeners consistently reproduced the same responses on the second occasion. This could be used to suggest that listeners have no difficulty in participating in a pitch perception task where no f_0 information is provided. It also suggests that there is a clear association between the formant frequencies and the perceived pitches of the whispered vowels (Thomas 1969).

A more recent general study asserts that listeners are able to perform voice recognition tasks by comparing a whispered vowel with voiced speech from a variety of speakers (Tartter 1991). Tartter makes no particular claim for which prosodic cue is possibly present during a whispered token. However, subjects consistently mistook tense and lax vowels for one another (eg. [I] vs. [i]; [U] vs. [u]) that differed in information present in the F1/F2 space.

From the studies outlined above, one might conclude that f_0 is not a necessary cue to the perception of pitch in whispered speech. The F1/F2 space seems to be an important and consistent finding amongst the reports summarised here. However, this issue lies outside the scope of the present paper and for future work to pursue.

3. Whisper in Norwegian

The aforementioned work conducted by Fintoft (1970) reported on a focused study on the dynamics of whisper and perception in Norwegian speech. His study was largely focused on the variance between dialects in voiced speech (an issue that is not pursued in the present study), but he also addressed the same issues with respect to whisper. To test a listener's ability to distinguish between voiced and whispered speech, Fintoft presented subjects with the pair *live* ('(a)live') – *livet* ('the life'), without any surrounding syntactic or semantic context. Subjects were able to make a reliable distinction, though the ability varied across dialect groups. Speakers of East Norwegian from the Trondheim area had the most difficulty in perceiving the difference, though by and large were able to complete the task in a satisfactory manner. Fintoft finds that listeners are likely to rely on other parameters apart from f_0. Intensity and syllable duration are potential candidate cues to the pitch accent distinction in whispered Norwegian speech.

Findings from Hadding-Koch (1961, 1962) corroborate the claim that intensity may be one such parameter with observations made from Swedish. She observed that subjects' judgements for two evenly spoken syllables in a continuous fashion with no pauses corresponded to pitch accent one. On the other hand, two evenly spoken syllables that were produced with low intensity or a small pause in between were often judged as pitch accent two.

Though Fintoft presents a thorough investigation into the potential prosodic cues for whispered pitch accent, there are some issues regarding his experiments that merit attention today. First, his study only tested subjects for accent perception on single-word utterances. The difficulty stemming from this is that a subject is given a single task, accent perception, on which all attention may be focussed. If this were avoided, by inserting the crucial words in an otherwise identical context, or by telling subjects to identify non-related differences in meaning, one is more certain to achieve results that are realistic, and speech-like. The second more important factor is the technical development that has taken place since Fintoft's work. With improved equipment and updated methods, such as speech synthesizers, we expect that the difference in importance between

context and phonetic cues, and the issue of phonetic cues themselves, may be addressed with a hope of stronger conclusions. Fintoft himself proposed that future work should incorporate synthetic stimuli into the whisper experiment so as to compare subject reactions to natural speech. We therefore set forth to revisit Fintoft's findings with respect to these issues.

4. The Experiment

Our focus in this paper, then, is the status of phonetic cues in the perception of whispered speech. In order to investigate this question, we devised a perceptual listening experiment. This test was designed principally to investigate whether pitch accent can be perceived in whispered speech without the assistance of context, i.e. just on the basis of phonetic cues. As has been shown previously, there has long been a disagreement and uncertainty on this issue, and empirical research presented from previous work has not addressed this issue from the perspective of a real-life communication situation. This feature makes them less interesting, for the reasons mentioned above.

Our experiment presented seven pairs of incomplete utterances, coupled with two possible completions. For each incomplete utterance, there was one critical word, occurring at the end of the utterance, which provided the pitch information necessary to obtain the only possible semantic disambiguation. Only incomplete utterances were included in the auditory stimuli. Our subjects were instructed to select one of two possible completions for each stimuli, and their choice of completion would be determined by their interpretation of the word with pitch accent. For example, in (1), the incomplete utterance is phonetically identical for both senses of the word (despite the orthographic difference). The intonation contour on the final word would be the only cue available to help disambiguation and the listener would choose the appropriate completion, either a) or b) based on which contour they think they have heard.

(1) Jeg hørte at noen bønder/bønner…
I heard that some farmers/beans…
'I heard that some farmers/beans…'
a. har reist til Oslo for å protestere.
have travelled to Oslo for to-INF protest.
'have travelled to Oslo to protest'
b. har blitt trukket tilbake av Rema fordi de var forgiftet.
have been drawn back by Rema because they were poisoned.
'have been withdrawn by Rema because they were harmful.'

The two possible completions in (1a) and (1b) are designed so that neither is felicitous with more than one pitch contour. In this example, it is evident that only farmers (accent 1 variant) can travel to Oslo to protest, whereas it is only beans (accent 2 variant) that can be withdrawn because of harmfulness. A complete list of the utterances used can be found in the appendix.

In order to address the issue of perception during whisper, we devised three test conditions for our stimuli. The conditions were designed to test a listener's ability in perceiving pitch accent in three separate acoustic environments. The first condition, the baseline condition, consisted of normal speech, recorded in a normal reading voice by a native Norwegian speaker (the second author) in a sound-proofed recording studio. The second condition had the same 7 utterance pairs, only this time recorded in a whisper. The third condition was resynthesised from the original voiced baseline condition. This resynthesis was performed with an application available in the Oxford Phonetics Laboratory, which splits the speech signal of its input, removes voicing information, and recombines spectral information and frication, to produce a 'mechanical' whisper. The purpose of the resynthesised condition was to show beyond a reasonable doubt whether there were additional phonetic cues involved in whispered speech which may assist a speaker to distinguish one pitch accent from another. The implementation of the third condition, if subjects proved able to correctly distinguish pitch accents 1 and 2, allows us to say with certainty that any extra cues for pitch accent - beyond pitch information - must be present already in normal speech.

For each of the three conditions, the seven pairs were randomised and repeated five times, giving the subjects 70 utterances to score for each condition. There were nine subjects, all native speakers of an East Norwegian dialect. The stimuli were presented in a sound-proofed room via loudspeakers in randomised order, and the subjects were asked to indicate which completion they preferred. For this purpose they were provided with a written text of all the stimuli, approximately like in (1) above, and could follow along and select the appropriate completions.

Four subjects completed the test in the Phonetics Laboratory, and five subjects did the test online[3]. Because of the nature of this study, and in order to secure similar test conditions for all subjects, we had to select and monitor closely who participated in the online part of the study. Detailed instructions were provided for the subjects doing the test online with regards to achieving the best results, and for preparing them for what to

[3] Available at http://teig.nvg.org/lyttetest/

expect. We found that performing the listening test online was a very helpful way of enabling subjects to participate who were not able to come to the laboratory. Even though this is not the traditional way of performing perception tests, it is becoming more common as access to internet technology is more widely available (cf. e.g. Caspers 2000).

5. Findings and discussion

The results show, as could be expected, an overwhelmingly correct disambiguation of the meaning in the incomplete utterances. Hardly any errors occurred in the responses to the baseline condition stimuli, with as much as 97 per cent correct identifications. On the other hand, the two other test conditions show strikingly similar disambiguation results. In natural whisper, the correct utterance completion was identified in 61 per cent of cases, and only a triflingly small difference exists between this and resynthesised whisper, which was identified correctly in 57 per cent of cases. Although one might look suspiciously at the absolute values in these results, we can already see emerging a close link between the two whispered conditions.

However, to provide answers to some of our questions, it is necessary to investigate whether subjects can actually be said to discriminate correctly between the two pitch accents, and to see whether the condition variable influenced the degree of correct perception of the intended meaning. To this end we have used a d' test, which is based on the z-score of the normal distribution and measures the effect of the independent variable on the dependent variable. This test shows that subjects were able to make the correct identifications in both whispered conditions, with $d' = 2.02$ for natural whisper, and $d' = 2.8$ for resynthesised whisper ($p < 0.05$, with $p = 0.05$ when $d' = 1.96$). As was mentioned above, this is an issue about which previous research (including Fintoft) does not provide firm conclusions. But with these results we can confirm that speakers of East Norwegian do discriminate between tokens that are traditionally seen as differing only in pitch accent, even when pitch is demonstrably absent. There is admittedly an effect on the level of accuracy at which discrimination is achieved, and a Student's t-test showed that discrimination was better in voiced speech than in natural whisper ($p < 0.05$), and better in natural whisper than in resynthesised ($p < 0.001$). This reflects the significance of the percentages mentioned above.

We have now confirmed with statistical evidence the hitherto intuitive claim that speakers of East Norwegian do not need to rely on context to disambiguate pitch accent in the absence of pitch information, that is in whispered speech. The statistical significance of our results leads to a

further conclusion that there must be other phonetic cues present in the speech signal which are utilised for disambiguating between two possible lexical meanings in place of prosodic information. This is an important finding, and should guide further research in this field. In connection with this study we made only preliminary investigations into the question of what cues are utilised, and no results can be confirmed. However, preliminary tests tentatively show there to be some support for syllable duration as one possible cue.

Possibly the most interesting finding of this study was identified in the d' test above, and is related to the results regarding the resynthesised whisper. We saw that the correct identification rate for these stimuli was significant, and even more strongly significant than the results for the naturally whispered stimuli. This result suggests that we should not expect to find that speakers are dependent on *additional* cues in whispered speech that are not also present in voiced speech. There is simply no evidence for the dependence of such extra cues. Moreover, if these extra cues had been present and actively utilised in utterance interpretation, the perception of pitch accent differences in resynthesised whisper should be impossible. On the contrary, this is fully possible, demonstrating that there is sufficient information in voiced speech. This is obviously not evidence against the presence of extra cues in whispered speech. What we have shown is that speakers do not depend crucially on such cues.

6. Conclusion

This paper has addressed the issue of speaker reliance on phonetic cues for utterance disambiguation when otherwise necessary pitch information is absent, such as during whispered speech. Results from a perception test of speakers of East Norwegian show conclusively that speakers can discriminate between pairs of words traditionally described in terms of purely prosodic difference, also in the absence of prosodic information. This suggests that there may be phonetic cues present in the speech signal which speakers may rely upon. Furthermore, tests using resynthesised whisper, demonstrate that speakers do not require extra phonetic cues present in whispered speech to compensate for the loss of pitch information. Rather, adequate cues for correct interpretation during whisper must be present in the voiced speech signal.

Appendix

Stimuli presented in the perception test. Each first line represents an incomplete utterance, and the two following lines represent the possible completions.

La oss håpe at gangen...
Let us hope that the-corridor/the-gait...
'Let us hope that the corridor/the gait…'
er stor nok til at vi kan henge fra oss frakkene der.
is big enough to that we can hang from us the-coats there.
'is big enough for us to leave our coats there.'
avslører tyven paa videoopptaket.
exposes the-thief on the-CCTV-recording.
'will reveal the identity of the thief on the CCTV.'

Du må fortelle meg om tanken…
You must tell me if the-tank/the-thought...
'Could you tell me whether the tank/the thought…'
kom plutselig, eller om du hadde vurdert det lenge.
came suddenly, or if you had considered it long.
'came suddenly, or whether you had been considering it for a while.'
kan fylles med vanlig bensin.
can fill-PASS with ordinary petrol.
'can be filled with ordinary petrol.'

Det er helt sikkert at [1]hakket/[2]hakke…
It is wholly certain that the-cut/pick-axe...
'It is certain that the cut/a pick-axe…'
ble laget med en skarp gjenstand.
'was made with a sharp object.'
bør være i enhver brevandrers oppakning.
should be in every glacier-walker's provisions.
'should be part of the luggage of everyone crossing on a glacier.'

Jeg håper noen finner…
I hope someone/some finds/Fins...
'I hope that someone will find…/some Fins…'
kan ta med litt vodka til festen.
can take with little vodka to the-party.
'can bring some vodka to the party.'
klokka som forsvant.
the-watch that disappeared.

Turistbrosjyra ga oss informasjon om [1]faret/[2]fare…
The-tourist-brochure gave us information about the-track/danger…
'The leaflet gave information about the track/the dangers…'
forbundet med fjellklatring.
connected with mountaineering.
som gikk mellom Tydal og Storerikvollen.
that went between Tydal and Storerikvollen.
'connecting Tydal and Storerikvollen.'

Jeg hørte at noen [1]bønder/[2]bønner…
I heard that some farmers/beans…
'I heard that some farmers/beans…'
har reist til Oslo for å protestere.
have travelled to Oslo for to-INF protest.
'have gone to Oslo to protest.'
har blitt trukket tilbake av Rema fordi de var forgiftet.
have been drawn back by Rema because they were poisoned.
'have been withdrawn by Rema because they were harmful.'

Vi er helt avhengige av at en eller annen skriver…
We are wholly dependent of that one or other writes/writer…
'We are fully dependent on someone writing/that some registrar…'
er tilstede under seremonien.
is present under the-ceremony.
'being present during the ceremony.'
et brev om dette til avisa.
a letter about this to the-newspaper.
'a letter about this to the newspaper.'

References:

Caspers, Johanneke. 2000. 'Experiments on the meaning of four types of single-accent intonation patterns in Dutch,' *Language and Speech* 43, 127-161.

Catford, James. 1977. *Fundamental Problems in Phonetics*, Edinburgh University Press, Edinburgh.

Fintoft, Knut. 1970. *Acoustical Analysis and Perception of Tonemes in Some Norwegian Dialects*, Universitetsforlaget, Oslo.

Hadding-Koch, Kerstin. 1961. *Acoustico-Phonetic Studies in the Intonation of Southern Swedish*, Gleerup, Lund.

Hadding-Koch, Kerstin. 1962. 'Notes on Swedish Word Tones,' *Proceedings of the International Conference of Phonetic Sciences, Helsinki*, Mouton, The Hague, pp. 630-638.

Ladefoged, Peter. 2001. *A Course in Phonetics,* 4th ed, Harcourt Brace, New York.

Lass, Norman J. 1996. *Principles of Experimental Phonetics*, Mosby Inc., St Louis, MO.

Meyer-Eppler, Werner. 1957. 'Realization of Prosodic Features in Whispered Speech,' *Journal of the Acoustical Society of America* 29, 104-106.

Nilsen, Randi Alice. 1992. *Intonasjon i interaksjon: sentrale spørsmål i norsk intonologi,* unpublished dr.art. dissertation, University of Trondheim, Norway.

Saunders, Ross. 2002. Online materials for the phonetics course at the Simon Fraser University, Burnaby, BC, Canada, http://www.sfu.ca/~saunders/l33098/L4/L4_6.html

Tartter, Vivian. 1991. 'Identifiability of vowels speakers of whispered syllables,' *Perception and Psychophysics* 49, 365-372.

Thomas, I.B. 1969. 'Perceived pitch of whispered vowels,' *Journal of the Acoustical Society of America* 46, 468-470.