# Using Prosodic Clues to Decide When to Produce Back-channel Utterances

**Nigel WARD**
Mech-Info Engineering, University of Tokyo,
Bunkyo-ku Tokyo 113 JAPAN
e-mail: nigel@sanpo.t.u-tokyo.ac.jp

**Motivation:** Real-time back-channel utterance production is needed for systems capable of natural conversation.

**Hypothesis:** We conjectured that it would be possible for Japanese to decide when to produce many back-channel utterances based on prosodic clues alone, without reference to meaning.

**Data:** We recorded a total of 47 minutes of short Japanese conversations between a 12 pairs of male university students, and labeled the 512 back-channel utterances (hereafter called "aizuchi").

**Analysis:** Neither vowel lengthening, volume changes, nor energy level (to detect when the other finished speaking) were by themselves good predictors of when to produce an aizuchi. The best predictor was a low pitch level.

**Implementation:** Our current best rule is: Upon detection of the end of a region of pitch less than .9 times the local median pitch and continuing for 150ms, coming after at least 600ms of speech, the system predicted an aizuchi 200ms to 300ms later, providing it had not done so within the preceding 1 second.

**Evaluation:** Predictions were scored as correct if the predicted aizuchi initiation point was within 500ms of that of an original human aizuchi. The above decision rule correctly predicted 38the aizuchis in the corpus, with an accuracy of 32predictions / 607 total predictions). For comparison, coverage and accuracy were 29in response to pre-recorded tracks (2 subjects, 5 minutes each).

**Post hoc analysis revealed:** Many of the missed aizuchis express not just "I'm listening, please go on", but also surprise, sympathy, confusion, disagreement etc. Scores for responses to non-Tokyo speakers were generally lower. Individual variation was very great. Many of the incorrect predictions were not unnatural. Accuracy was highest when the speaker was engaged in a longish explanation.

**System:** We also built a real-time system based on the above decision rule. A human stooge steered the conversation to a suitable topic and then switched on the system. After switch-on the stooge's utterances and the system's outputs, mixed together, produced one side of the conversation. After switch-on the stooge produced an average of 5 turns (including questions, answers, and aizuchi) and the system contributed an average of 5.2 turns (all aizuchis, specifically, neutral grunts, of course), over the course of a minute. Debriefing revealed that none of the 4 subjects had realized that his conversation partner had become partially automated.

**Future plans:** We intend to do all the right things.

**Gratuitous Table:** Here are the aizuchi we observed, and for each, the fraction of the observed aizuchi which the above rule was able to account for. Of course, the transliteration leaves much to be desired. This table suggests many hypotheses, some of which we intend to investigate.

0.0 = 0/ 1 a:a-maji
0.0 = 0/ 1 a:aa-asagaya-aa-wakaru-wak
0.0 = 0/ 1 a:aa-ayase-hatsu-ne
0.0 = 0/ 1 a:aa-hontoo
0.0 = 0/ 1 a:aa-maa-nee
0.0 = 0/ 1 a:aa-naruhodo-nee
0.0 = 0/ 1 a:aa-sengenjaya-ka-uun
0.0 = 0/ 1 a:aa-sorewa-nee
0.0 = 0/ 1 a:aa-yappari
0.0 = 0/ 1 a:aaaaa
0.0 = 0/ 1 a:aaaaaa
0.0 = 0/ 1 a:aaaaaaaa
0.0 = 0/ 1 a:aasoodakke
0.0 = 0/ 1 a:achigaundesuka
0.0 = 0/ 1 a:ahaa
0.0 = 0/ 1 a:aheheha
0.0 = 0/ 1 a:ahontoo
0.0 = 0/ 1 a:anaruhodo
0.0 = 0/ 1 a:ano-baai-no-kazu-o-iu-tt
0.0 = 0/ 1 a:asa
0.0 = 0/ 1 a:ashita-tanjoobi-nano
0.0 = 0/ 1 a:berunuui?
0.0 = 0/ 1 a:cough
0.0 = 0/ 1 a:da-dame
0.0 = 0/ 1 a:doitsugo-ka
0.0 = 0/ 1 a:doko-no?
0.0 = 0/ 1 a:e
0.0 = 0/ 1 a:ee?
0.0 = 0/ 1 a:ee?-heh
0.0 = 0/ 1 a:ehehe
0.0 = 0/ 1 a:fuji
0.0 = 0/ 1 a:gentei-ka
0.0 = 0/ 1 a:ginza?
0.0 = 0/ 1 a:ha
0.0 = 0/ 1 a:haaa
0.0 = 0/ 1 a:hahaheha
0.0 = 0/ 1 a:he?
0.0 = 0/ 1 a:hee-hehe
0.0 = 0/ 1 a:hehaha
0.0 = 0/ 1 a:hehuhuhahaha
0.0 = 0/ 1 a:hh
0.0 = 0/ 1 a:hiru?
0.0 = 0/ 1 a:hm-um-naruhodo
0.0 = 0/ 1 a:hmhaha
0.0 = 0/ 1 a:hmhmm
0.0 = 0/ 1 a:hmmm
0.0 = 0/ 1 a:mmm
0.0 = 0/ 1 a:naa-hehe
0.0 = 0/ 1 a:nanda
0.0 = 0/ 1 a:nani
0.0 = 0/ 1 a:nani?
0.0 = 0/ 1 a:naruh
0.0 = 0/ 1 a:nee
0.0 = 0/ 1 a:nezu-eki
0.0 = 0/ 1 a:nifun
0.0 = 0/ 1 a:noru?
0.0 = 0/ 1 a:oaa
0.0 = 0/ 1 a:oh
0.0 = 0/ 1 a:shimamura?
0.0 = 0/ 1 a:so-atarii
0.0 = 0/ 1 a:sokasoka
0.0 = 0/ 1 a:soo-naa
0.0 = 0/ 1 a:soo?
0.0 = 0/ 1 a:soodanee
0.0 = 0/ 1 a:soodesu
0.0 = 0/ 1 a:soojanaino?
0.0 = 0/ 1 a:sookakaa
0.0 = 0/ 1 a:soonano
0.0 = 0/ 1 a:soosoosoosoosoo
0.0 = 0/ 1 a:sorewanai?
0.0 = 0/ 1 a:sorewasooda
0.0 = 0/ 1 a:tariru-nda
0.0 = 0/ 1 a:tonari
0.0 = 0/ 1 a:uhuhu
0.0 = 0/ 1 a:unhaihai
0.0 = 0/ 1 a:unun
0.0 = 0/ 1 a:usoo
0.0 = 0/ 1 a:uu
0.0 = 0/ 1 a:uuhn
0.0 = 0/ 1 a:ya
0.0 = 0/ 1 a:yaa-maa-nee
0.0 = 0/ 1 a:yoku-wakaran-nee
0.0 = 0/ 2 a:aaaa
0.0 = 0/ 2 a:ehaha
0.0 = 0/ 2 a:hehe
0.0 = 0/ 2 a:hontoo
0.0 = 0/ 2 a:ka
0.0 = 0/ 2 a:ska
0.0 = 0/ 2 a:sooka
0.0 = 0/ 2 a:sookaa
0.0 = 0/ 2 a:uunn
0.0 = 0/ 3 a:ahontoni

0.0 = 0/ 1 a:hnhaa
0.0 = 0/ 1 a:hnhn
0.0 = 0/ 1 a:hoh
0.0 = 0/ 1 a:hoo
0.0 = 0/ 1 a:hooon
0.0 = 0/ 1 a:huhuhu
0.0 = 0/ 1 a:hunaha
0.0 = 0/ 1 a:ie-de-nedotta
0.0 = 0/ 1 a:ii-des-nee
0.0 = 0/ 1 a:ii-na
0.0 = 0/ 1 a:inhale
0.0 = 0/ 1 a:iyadane
0.0 = 0/ 1 a:jikken
0.0 = 0/ 1 a:jimichi-tte-heh
0.0 = 0/ 1 a:joohoo?
0.0 = 0/ 1 a:koomuin
0.0 = 0/ 1 a:kowai-ne
0.0 = 0/ 1 a:kumiawaseron?
0.0 = 0/ 1 a:maa-neh
0.0 = 0/ 1 a:mechacha-kakate-masu-ne
37.5 = 3/ 8 a:uh
40.0 = 2/ 5 a:asoonanda
42.9 = 6/14 a:unn
50.0 = 1/ 2 a:eee
50.0 = 1/ 2 a:hahahaha
50.0 = 1/ 2 a:huun
50.0 = 1/ 2 a:soodayone
50.0 = 1/ 2 a:soonanda
0.0 = 0/ 4 a:hee
0.0 = 0/ 6 a:hmm
11.1 = 1/ 9 a:naruhodo
12.5 = 1/ 8 a:haha
16.7 = 1/ 6 a:haihai
18.2 = 2/11 a:soosoo
19.2 = 5/26 a:aa
20.0 = 1/ 5 a:hm
20.0 = 1/ 5 a:hontoni
25.0 = 1/ 4 a:he
25.0 = 1/ 4 a:soo
28.6 = 2/ 7 a:uuh
33.3 = 1/ 3 a:aaa
33.3 = 1/ 3 a:asoo
33.3 = 1/ 3 a:asooka
33.3 = 1/ 3 a:ee
33.3 = 1/ 3 a:nanto
33.3 = 1/ 3 a:ooaa
33.3 = 1/ 3 a:soosoosoo
33.3 = 2/ 6 a:mm
100.0 = 1/ 1 a:uuaa
100.0 = 1/ 1 a:uum
100.0 = 1/ 1 a:yappa-minna-soonanda
100.0 = 1/ 1 a:yarimakuten-no
100.0 = 2/ 2 a:aa-haihai
100.0 = 2/ 2 a:huh
100.0 = 2/ 2 a:hun
100.0 = 2/ 2 a:ooa

50.0 = 1/ 2 a:soosoosoosoo
50.0 = 1/ 2 a:uh?
50.0 = 1/ 2 a:uhh
50.0 = 1/ 2 a:um-maa-nee
50.0 = 1/ 2 a:un?
50.0 = 1/ 2 a:uun
50.0 = 2/ 4 a:hahaha
53.5 = 46/86 a:un
66.7 = 2/ 3 a:hunhun
75.0 = 3/ 4 a:umm
100.0 = 1/ 1 a:a-mina-kore-sonanda
100.0 = 1/ 1 a:aa-maa-soo-dakedo
100.0 = 1/ 1 a:aa-nanto
100.0 = 1/ 1 a:aa-unn
100.0 = 1/ 1 a:aa-zutto-tokyo
100.0 = 1/ 1 a:ah-nandemonai

100.0 = 1/ 1 a:aha-haihai
100.0 = 1/ 1 a:ahahaihai
100.0 = 1/ 1 a:ahon
100.0 = 1/ 1 a:asoo-hahaha
100.0 = 1/ 1 a:asoonan
100.0 = 1/ 1 a:asooo
100.0 = 1/ 1 a:damearee
100.0 = 1/ 1 a:ee-ichioo
100.0 = 1/ 1 a:famiconkei
100.0 = 1/ 1 a:haa
100.0 = 1/ 1 a:haa-ha
100.0 = 1/ 1 a:hai
100.0 = 1/ 1 a:hai-nanisen?
100.0 = 1/ 1 a:hai?
100.0 = 1/ 1 a:haihaihai
100.0 = 1/ 1 a:heh
100.0 = 1/ 1 a:hehehe
100.0 = 1/ 1 a:heru-ne
100.0 = 1/ 1 a:ho
100.0 = 1/ 1 a:hoho
100.0 = 1/ 1 a:huha
100.0 = 1/ 1 a:hun-yappa-ne
100.0 = 1/ 1 a:huuun
100.0 = 1/ 1 a:iiemasu-ne
100.0 = 1/ 1 a:inhale-huun
100.0 = 1/ 1 a:joohoo
100.0 = 1/ 1 a:kettei-shite-iru-nda
100.0 = 1/ 1 a:kinishite-inai-no?
100.0 = 1/ 1 a:kirai-nano
100.0 = 1/ 1 a:konai-no?
100.0 = 1/ 1 a:mita-sen?
100.0 = 1/ 1 a:nagasaki
100.0 = 1/ 1 a:nn
100.0 = 1/ 1 a:oo
100.0 = 1/ 1 a:shooganai-ndesuka
100.0 = 1/ 1 a:sniff
100.0 = 1/ 1 a:soodane
100.0 = 1/ 1 a:sorekara?
100.0 = 1/ 1 a:suriee
100.0 = 1/ 1 a:un-sore-wa-aru-ne
100.0 = 1/ 1 a:unununun
100.0 = 1/ 1 a:uso

Keywords: reflex, back-channel, prosody, dialog, subsumption, dialog