

# How do the elderly talk to a natural language call routing system?

*Knut Kvale, Ragnhild Halvorsrud*

Telenor Research & Innovation, Snarøyveien 30, N-1331 Fornebu, Norway

{Knut.Kvale, Ragnhild.Halvorsrud}@telenor.com

## Abstract

This paper investigates spontaneous utterances in response to a telecom IVR prompting the caller with “*Please tell me in a few words what your inquiry is about*”. In responses from a typical caller base (n=29830, age-mixed) the median utterance length was 3 words. A qualitative analysis (n=74) of senior speakers (age 70+) and adult speakers (age 50 and below) revealed that seniors were significantly more verbose (median of 11 and 3 words per utterance, respectively). The seniors generally expressed themselves more formally than those below 50. Our findings suggest that some seniors confuse the IVR with a telephone answering machine.

**Index Terms:** natural language, age, call routing, verbosity

## 1. Introduction

Traditional interactive voice response (IVR) systems force callers to listen to a series of voice prompts to determine an appropriate response. On the other hand, natural language understanding (NLU) technology enables speech-based IVR-systems to understand a huge range of utterances. These open dialog systems prompt the users in an open manner, letting the users describe their needs in their own words - which makes the human-machine interaction more natural.

In the last decade many telecom operators have gradually replaced system-driven dialogs with open dialogs in their public customer care services, e.g. AT&T Consumer Services' How May I Help You? (HMIHY) [1,2], France Telecom [3], TeliaSonera in Sweden [4], and Telenor IVR 05000 in Norway [5].

In this paper we present a study of how people actually talk to an NLU-based call routing system called IVR 05000. Section 2 gives a brief overview of the system, and section 3 describes how the speech data sets were extracted. In section 4 the results of our analyses are presented.

## 2. System overview

### 2.1. Telenor IVR 05000

IVR 05000 is Telenor's public customer care service for fixed and internet/broadband issues in Norway available on phone number 05000. Telenor Customer Service introduced speech enabled IVR in their fault report center in July 2001 and in the IVR 05000 service in November 2002. The dialogs at that time were system-driven, prompting the callers with simple “yes” or “no” questions, or with questions with limited response alternatives (e.g. “private” or “enterprise”). Although the service was simple with respect to automatic speech recognition (ASR) technology, it paved the way for Norwegians to get used to speaking to a machine.

Since then, services and spoken language technology have developed rapidly. In the period from December 2006 to February 2007 Telenor gradually introduced open dialog in

their IVR 05000. Based on the callers' free description of their needs the system carries out automatic speech recognition and a semantic analysis which assigns so-called semantic tags. Based on this semantic categorization and dialogs with the callers, the system (ideally) routes the calls to one of totally 46 different self-service categories or “destinations” or to a customer agent. Each of the self-service categories has their own dialogs for helping the users solve their problems or get relevant information.

IVR 05000 handles around 25000 calls per day. It runs on the Dolphin HotVoice platform [6] which has been integrated with the Nuance® OpenSpeech™ Recognizer with the so-called SpeakFreely™ feature [7]. As no Norwegian text-to-speech (TTS) synthesis has yet been regarded as good enough for this service, the system prompts are recorded female speech.

### 2.2. The greeting dialog

In this paper we have analyzed recordings from July and September 2007. At that time all callers to IVR 05000 were welcomed with the following prompt (translated from Norwegian): “*Welcome to Telenor. Please enter the relevant telephone number for your inquiry. If you are a private customer without a landline connection, please type in your mobile phone number. If you are not a Telenor customer, please press the hash key*”. The next stage in the interaction depends on the caller's actions to this prompt – whether he types in a number, presses the hash key, takes no action at all, or makes other types of input. If the caller types a proper telephone number, the system acknowledges the input with the following: “*Thanks. I will retrieve some information. Just a moment please*”. The system uses the given eight-digit input to retrieve information from a number of customer management systems about subscription type, age of subscriber, previous inquiries, reported errors in the area etc. After a brief pause the system prompts the caller with: “*Please tell me in a few words what your inquiry is about*”. The callers responses to this last prompt form the basis of our analysis in this paper.

## 3. Method

Two data sources, a large data set and a small data set, have been analyzed in this study. The large data set contained 29830 transcribed utterances (n=29830), based on calls to IVR 05000 collected in July 2007. This large data set is representative for Telenor's customer base, but this material contained no information about the callers' age or gender.

In order to search for age characteristics, utterances associated with subscribers aged 70 and over were selected for analysis, forming a smaller, but more comprehensive set (n=124). Here we used recordings from September 2007. The following subsections provide details about sample selection, age determination and analysis of this small data set.

### 3.1. Sample selection

IVR 05000 automatically assigns an age tag when the caller keys in a number (see section 2.2). However, since a fixed line telephone may serve a household consisting of several members, it is not granted that the caller always is the subscriber. Furthermore, senior citizens often receive assistance from younger family members, friends, and others when interacting with automated services (see section 5.2). Nevertheless, being able to identify the caller as the subscriber eliminated any uncertainty about age, and was used as an approach in combination with pure voice-based age determination.

Let  $H_0$  denote the hypothesis "the caller is the subscriber". Gender determination based on voice is known from the literature to be quite precise [8] and was an immediate way to falsify  $H_0$ . A call placed from a mobile phone with the same subscriber as the number keyed in supported  $H_0$  since a mobile phone is often regarded as a personal device in Norway. Sometimes the content of the utterance itself turned out to be very useful. The use of pronouns as, e.g. "I want to change my subscription" or "My dialing tone is missing" supports  $H_0$ . Finally, some utterances could be decisive in themselves, e.g. "my name is NN" or "my brother's phone". An overall evaluation based on these clues was often sufficient to falsify or verify  $H_0$ .

### 3.2. Age determination

Several previous reports on age determination have shown that listeners are able to make fairly accurate judgments about speakers' age from voice cues only. For instance, a group of listeners were able to assign voices of two age-groups, over 65 and under 35, to their correct age-group with very high accuracy [9]. It is stated that "...when subjects read aloud the same written passage, age can usually be told within ten years" [9]. This statement has been confirmed in later tests [8, 10]. Common to these reports is the use of read-aloud speech for age determination. Contrary to this, the present material is based on spontaneous speech. This is a far more realistic context, providing more perceptual cues for determining the speaker's age, such as articulation rate, breath sounds, degree of hesitation and stuttering effects [11], and finally the occurrence of old-fashioned expressions, words or grammar.

### 3.3. Classification in age groups

In the age determination procedure we chose general age descriptions rather than age groups, as recommended in [8]. The authors classified the voice recordings independently of each other, initially in three groups:

- adults (age 50 or under)
- middle-aged (age 51-69)
- seniors (age 70 and over)

Non-conclusive results between the authors were left out of the analysis ( $n=19$ ). The samples were then compared to the group already defined as seniors based on identifying the caller as the subscriber. From a total of 34 seniors, five cases were judged as belonging to the middle-aged group, leaving 29 cases where a correct age group was assigned. This shows that we underestimated some of the seniors, a tendency known from the literature [8].

Earlier reports show that listeners are able to differentiate voices of young adults from those of elderly speakers with impressive accuracy [9]. We therefore left out the middle-aged group ( $n=16$ ). Also, recordings lacking a response to the greeting prompt were discarded ( $n=15$ ).

From a total number of 124 recordings, 32 and 42 cases remained in the adults' and seniors' groups, respectively. The adults consisted of 20 females and 12 males. The seniors consisted of 17 females and 25 males.

### 3.4. Transcription and analysis

The utterances were transcribed orthographically. A cutoff was transcribed if a speaker did not complete the pronunciation of a word. If the cutoff sounded as a more or less complete word, it was counted as an extra word, e.g. "tele telephone" was regarded as two words. Stuttering with only one sound or an incomprehensible word was not counted more than once, e.g. "t t t telephone" was regarded as one word. That is, all speech sounds that were perceived as meaningful Norwegian words were counted as words.

In the telecom sector acronyms like ADSL are often used in product names. Although pronounced as separate letters, they were counted as one word.

In Norwegian there are three ways of pronouncing a numeral. The number 28 may be spoken either as three words: "eight-and-twenty" (the old counting style), as two words: "twenty-eight" (the new counting style), or digit by digit: "two eight" [12]. Numerals were counted in the way they were said by the speaker.

Utterance length was obtained by counting the number of words in an utterance. Note that all the counts were performed in the original language (Norwegian). Comparison of utterance lengths within the two age groups was done by using the Mann-Whitney U test (in SPSS 14.0.1.).

## 4. Results

### 4.1. The large data set

Figure 1 shows the distribution of utterance lengths in the large data set. Here, one-word responses were most frequent constituting 29.4% of all the utterances. The most frequent utterance was the single word "internet", and the second-most frequent response was "invoice". (Such responses are vague

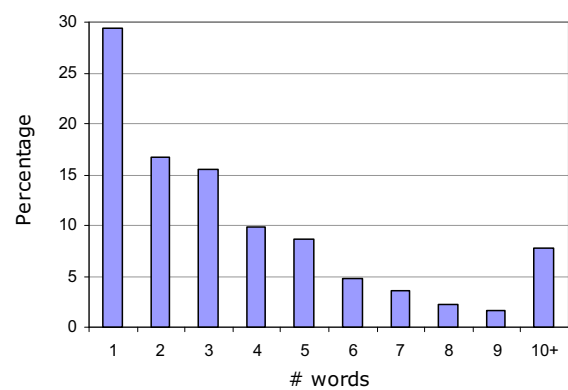


Figure 1: Distribution of utterance lengths in the large data set.

formulations that require a new dialog turn in the system to get the correct call routing.) 1311 detected cutoffs were counted as words. In addition there were 304 restarts, where the speaker typically repeated the whole word. Some of the callers told the IVR long life-stories. In these cases only the first part which described the intention of the inquiry was transcribed. This approach was applied to 302 utterances (about 1%) and led to a maximum utterance length of 43

words. Thus, the average utterance length may be slightly underestimated. Summing up, the large data set contained a total of 117133 words. The average utterance length was 3.9 words, and the median was 3 words.

#### 4.2. Detailed analysis of utterances

In the small data set we found that the adult callers typically responded with short utterances of a few words, forming a sentence without any pronouns, e.g. “Telephone doesn’t work” or just “Invoice” or “Moving”. Three-word responses were most frequent within the adult group, constituting about one third of all utterances.

Seniors typically expressed themselves using longer utterances, e.g. “The inquiry is about postponement of an invoice. It is the twentieth of September”. Seniors used pronouns more frequently (“my invoice”, “my Internet connection”), and they often parroted parts of the IVR prompt (“It is about...”, “My inquiry is about...”), see section 4.3.

Figure 2 shows a box plot of the utterance lengths within the two age groups. The distribution was skewed and contained outliers. The outlier in the senior group uttered 76 words (!) - almost telling a whole life story.

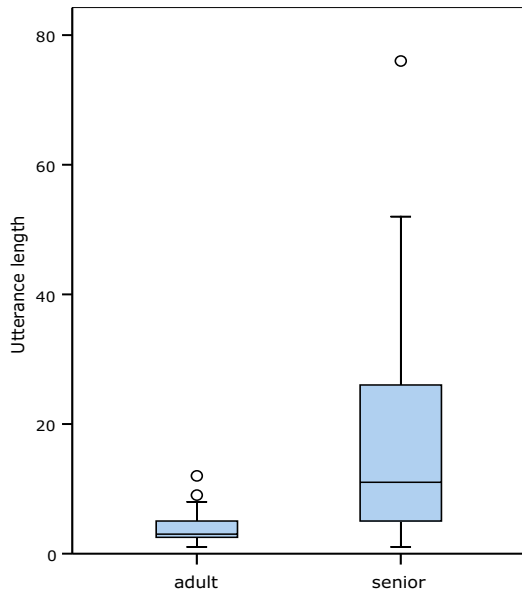


Figure 2: Box plot of utterance lengths among the two age groups. The horizontal bar corresponds to the medians.

The utterance length varied between 1 and 12 words for the adult group and 1 and 76 for the senior group. Median values were 3 and 11, respectively. Upper and lower quartiles can be found in Table 1. The difference in utterance length between adults and seniors was statistically significant ( $p < 0.001$ ). A small tendency towards longer utterances could be discerned among female callers, but this was not a significant result. The IVR system often interrupted long utterances leading to misunderstandings and confusions and problems with the barge-in functionality.

One utterance from a senior caller was very illustrative with respect to spontaneous speech effects (in Norwegian): “den regn henvendelsen gjelder regninga som æ ha fått æ kankje skjønne at det kan gå at det kan være rett over tolv hundre og tre og tredve krona og fire øre”. Verbatim translation into English with spontaneous speech effects in square brackets: “the invo[cutoff] [cough] the inquiry is

about the invoice I have received I can’t understand that it can go [pause] that it can be correct [pause] over twelve hundred and thirty kroner and four cents.” This caller was probably prepared to ask about her invoice, starting out with “the invo”, but seemed to be surprised by the prompt and changed her mind, stopped the pronunciation in the middle of the word and then parroted the last words of the system prompt. She then stated the main reason for the inquiry, pausing briefly, and continued with a detailed description of the problem. She talked in dialect without making any attempt at “normalizing” her speech. (In the end the caller was interrupted by the system which asked for more precise information).

Group	Size	Utterance length		
		range	median	quartiles
Adult (age < 50)	n=32	1-12	3	q1=2.25 q3= 5
Senior (age > 70)	n=42	1-76	11	q1=5 q3=26.25

Table 1: Characteristics of utterance lengths in the two age groups in the small data set.

#### 4.3. Parrotting

When interacting with an IVR some callers started their response by repeating the last few words of the system prompt. This phenomenon is referred to as parrotting, and the occurrence of parrotting can be found in Table 2. The greeting dialog in IVR 05000 ends with “Please tell me in a few words what your inquiry is about”. The word “inquiry” (“henvendelse” in Norwegian) is regarded as a formal and a bit old-fashioned word. It would be more natural to respond with “It’s about...” (“Det gjelder...”).

In the large data set, 1234 out of 29830 utterances (4.1%) began with “It’s about...” and only 232 (0.8%) said “My inquiry is about...”. In the small data set 5 of the 42 seniors started with “My inquiry is about...” and none of the adults used this expression. Furthermore, 8 seniors and 3 adults started with “It’s about...”. In addition, 3 seniors started with: “Yes, it’s about...”. Hence, there was a tendency towards more parrotting among the seniors.

Group	Parrotting	“I”	“We”	“My”	“Our”
Adult	3/32	1/32	-	-	-
Senior	16/42	22/42	-	6/42	-
Large data set	4.9%	16.3%	1.5%	2.3%	0.2%

Table 2: Occurrence of parrotting and pronouns.

#### 4.4. Personal wording

Some callers used a personal style with pronouns when talking to the IVR. Table 2 shows that the seniors used pronouns much more frequently than both the adult group and callers in the large data set. More than half of the seniors used the personal pronoun “I”, but only one adult did that. In comparison, 16.3% of the utterances in the large data set contained “I”. The plural form “we” was not found in the small set, but occurred in 1.5% of the utterances in the large data set. A similar pattern of use was found for the possessive pronoun “my”. Six of the seniors used this word, while none

of the adults did. In the large data set 2.3% of the utterances contained “my” and only a few utterances contained “our”.

In Norwegian the possessive pronoun can be placed before the noun e.g. “min telefon” (my telephone), or after the noun, “telefonen min”. The former is a more formal style than the latter. In the large data set two out of three speakers (66.2%) used the formal style.

We also observed a few cases in the large data set addressing the system as “you” (0.5%). A typical utterance was “*You may help me with...*”. Finally, 2 seniors and 25 callers in the large data set rounded off their statement saying “*thank you*”.

## 5. Discussion

### 5.1. Why are the elderly more verbose?

We found that the elderly were significantly more verbose than the adult group. There may be many reasons for this. One is to what extent they are used to speaking to IVR-systems in general, and to IVR 05000 in particular. Clearly, persons who have called the IVR 05000 several times before will learn how to speak to the system.

Although the analysis of how the seniors and adults talked to an IVR system was limited to 74 persons, we have found some interesting tendencies. Based on careful listening it seems as if some of the elderly have a mental model of IVR 05000 as an answering machine where they simply leave their message and assume that someone will listen to and respond to this later. These callers are rather verbose and may even round off with “thanks” before they hang up. Many of the elderly state their inquiry in many different ways in the same utterance. It seems as if they are uncertain whether they use correct terms for describing the technical problem or even the correct word for customer agent. These persons use many words and they hesitate a lot and may eventually be interrupted by the IVR due to time out.

As shown in section 4.3 some seniors tend to be more formal in their wording, stating their inquiry in full sentences instead of simply stating a few key words.

Some callers state that the inquiry is about a certain telephone number. We found that only the seniors applied the old counting style in such cases (4 cases in our material).

### 5.2. The trusted assistants

A single selection criterion, namely a subscriber age of 70+, was used for obtaining the sample constituting the small data set (n=124). The analysis revealed that only one third of these callers (n=42) were aged 70+. Evidently, many seniors received help from others. In order to figure out why they did not make these calls themselves, phone agents at Telenor's call center interviewed 48 of the helpers in October 2007 (identified and recruited by inquiries to IVR 05000). These interviews gave an indication of who the helpers were, and why their assistance was needed. Most of the trusted assistants were younger relatives of the elderly subscriber: About two thirds (67%) were either children or children-in-law. Grandchildren were also represented among the trusted assistants.

The main reasons for assisting the seniors were: Illness (n=15), “It is usually me who sort things out” (n=9), sight- or hearing problems (n=6), “It is too difficult” (n=4). Thus, illness or reduced operability among the seniors were the main reasons for using trusted assistants. Interestingly, only

four assistants considered IVR 05000 too complex for seniors.

## 6. Conclusions and future work

In this study we have investigated spontaneous utterances in response to a telecom IVR, prompting the caller with: “*Please tell me in a few words what your inquiry is about*”. We found that senior speakers (age 70+) were significantly more verbose than adult speakers (age 50 and below). Furthermore, the seniors used a more formal speaking style, used more personal pronouns, and were more likely to parrot the prompt. Our findings suggest that some seniors may confuse the IVR with a telephone answering machine. We recommend that verbosity as well as formal and informal speech style should be considered when designing dialogs and tuning parameters of timeout and barge-ins.

## 7. Acknowledgements

We would like to thank Kristian Schüffner, Cathrine Olsen, Gjermund Lia and Christian Nyegaard in Telenor's customer service department for all their help and for providing data. We also want to thank Narada Dilp Warakagoda, Kathrine Frey, Jon Emil Natvig and Arne Kjell Foldvik for their support.

## 8. References

- [1] Gorin, A. L., Riccardi, G. and Wright, J. H., “How May I Help You?”, *Speech Communication*, Vol. 23, pp.113-127, 1997.
- [2] Gorin, A., et. al., “Automated natural Spoken Dialog”, *IEEE Computer Magazine*, vol 35(4), pp. 51-56, 2002.
- [3] Dammati, G., et.al., “Spoken language understanding strategies on the france telecom 3000 voice agency corpus”, *IEEE ICASSP 2007*, pp. IV. 9-12. 2007.
- [4] Lövenlid, B. “Intelligent telefontjänst utvecklad i Uppsala”, *Uppsala Nya Tidning*, 2006-07-01. (In Swedish), Available from URL [http://www2.unt.se/avd/1,1786,MC=1-AV\\_ID=511475,00.html](http://www2.unt.se/avd/1,1786,MC=1-AV_ID=511475,00.html)
- [5] Nuance, “Telenor Finds Its Way with Nuance Call Steering”, Nuance press release September 19, 2007. Available from URL: [http://www.nuance.com/news/pressreleases/2007/20070919\\_telenor.asp](http://www.nuance.com/news/pressreleases/2007/20070919_telenor.asp)
- [6] Dolphin HotVoice. Available from URL: <http://www.dolphin.no/filestore/HotVoiceProductBrochure.pdf>
- [7] Nuance open speech recognizer. Available from URL <http://www.nuance.com/recognizer/openspeechrecognizer/>
- [8] Cerrato, L., Falcone, M. and Paoloni, A., “Subjective age estimation of telephonic voices”. *Speech Communication*, 31(2), pp. 107-112, 2000.
- [9] Ptacek, P. H. and Sander, E. K., “Age recognition from voice”. *Journal of Speech and Hearing Research*, 9(2), 273-277, 1966.
- [10] Neiman, G. S. and Applegate, J. A., “Accuracy of listener judgments of perceived age relative to chronological age in adults”. *Folia Phoniatrica*, 42(6), 327-330, 1990.
- [11] Shriberg, E., “Spontaneous Speech: How People Really Talk and Why Engineers Should Care”, *INTERSPEECH 2005*, pp.1781-1784, 2005.
- [12] Kvale, K. and Foldvik, A. K., ““Four-and-twenty, twenty-four”. What's in a number?”, *Proc. European Conference on Speech Communication and Technology (EUROSPEECH-97)*, vol.2, pp. 729-732, 1997.