
Tom Brøndsted & Erik Aaskoven

1 Department of Communication Technology, Aalborg University, Denmark
2 Efaktum, Hjørring, Denmark

Abstract

The public-funded project "Indtal" ("Speak-it") has succeeded in developing a Danish voice-controlled utility for internet browsing targeting motor-handicapped users having difficulties using a standard keyboard and/or a standard mouse. The project has been designed and implemented in collaboration with an advisory board of motor-handicapped (potential) end-users and underlies a number of a priori defined design criteria: learnability and memorability rather than naturalness, minimal need for maintenance after release, support for "all" web standards (not just HTML conforming to certain "recommendations"), independency of the language on the websites being browsed, etc. These criteria have lead to a primarily message-driven system interacting with an existing browser on the end users’ systems.

1. Introduction

The project Indtal ("Speak-it") has aimed at developing a Danish voice-controlled tool for internet browsing targeting motor-handicapped users having difficulties using a standard keyboard and mouse. The project was funded by the Danish National IT and Telecom Agency under the Ministry of Science and ran from primo January 2004 to ultimo February 2005. The project partners were the Department of Communication Technology at Aalborg University being in charge of the actual speech recognition technology and the software company Efaktum in Hjørring implementing the back-end communication between the recogniser and the browser engine of the system. Further the project involved two non-technical partners, "Specialskolen for Voksne, Vendsyssel" and "Teknologicentret for Handicappede, Nordjylland's Amt", institutions located in North Jutland and specialized in compensating courses and consultancy for adults with special needs, including adults with physical disabilities. These non-technical partners have been in charge of project dissemination, public relations, contact with an advisory-board of potential end-users, and the maintenance of a website www.indtal.dk where disabled users can download the browser for free.

The background of Indtal is of course the increasing focus on e-inclusion, accessibility for persons with disabilities, etc. This focus has for instance in the USA lead to ADA ("Americans with Disabilities Act") requirements being applied also to the web [1]. In Denmark and other countries there are lobbies eager to promote regulation similar to the American.

In spite of the new focus on accessibility it should be kept in mind that speech technology has "always" (at least since the mid 1990'ies) implicitly or explicitly addressed users with visual or mobile disabilities, sometimes "disguised" as a more general goal of enabling eyes-free/eyes-busy or hands-free/hands-busy access to web browsing and other applications.

Consequently, the Indtal-browser is far from something entirely innovative (apart from the fact that it supports Danish!). What separates the Indtal-browser from the similar systems supporting English or other (typically major) languages is rather a suite of design and implementation criteria to be outlined below.

2. Design & Implementation Criteria

The Indtal-browser differs from other "alternative web browsers" by a number of design and implementation criteria. We define an alternative web browser as a browser offering an alternative to either standard visual output rendering or to standard keyboard and mouse input control – or both. A number of such alternative browsers are listed at W3C's website hosting the Web Accessibility Initiative (WAI) [2] and similar websites devoted to disabled users. We distinguish two major groups both of which can deploy speech recognition and/or speech synthesis:

1) Group 1: The largest group explicitly addresses visually impaired users.
   a. Structured output is either sent to a Braille display or a speech synthesizer (or both).
   b. User input is given by means of a set of function keys or an equivalent set of spoken commands (but not by mouse that conflicts with visual disabilities).
   c. A relatively deep parsing of the structure and content of the web pages being browsed is deployed.
   d. Examples are Braillesurf [3][4], BrookesTalk [5], Emacsspeak [6], and Homer [7].

2) Group 2: Another group addresses the problem of enabling hands-free and (to some extent) eyes-free browsing.
   a. User input is given by spoken commands.
   b. Output is unaltered visual rendering enriched with so-called Saycons™ ("sayable icons" informing the user how to activate the various links by voice) and optionally combined with speech synthesis.
   c. Unless much of the functionality characterizing the first group is incorporated, no deep parsing of web-pages is deployed.
   d. Examples are Conversa Voice Surfer (formally Conversa Web) [7], HFB (HandsFree Browser...
by EduMedia) and add-ons shipped with certain versions of IBM’s ViaVoice and Dragon Natural-Speaking.

The Indtal-browser belongs to the latter group though it has an explicit focus on end-users with mobile disabilities. Generally web-content is highly visually oriented and it makes in our view no sense to attempt to support eyes-free browsing unless the needs of visually impaired users are addressed explicitly. Combining hands-free and eyes-free facilities for web browsing hardly makes sense at all. Users with both mobile and visual handicaps are extremely few and their most severe every-day problems do not encompass access to the web.

Apart from the stricter focus on the end-users, the Indtal-browser has been implemented to meet the following five criteria:

2.1. Minimizing Future Maintenance Requirements

A quick inspection of the “alternative web browsers” listed on the WAI site and similar websites devoted disabled users reveals that many browsers have been quietly withdrawn leaving no trace except for the broken links on the referring sites. In few cases the systems are still available for download, but have been “frozen” and are not maintained anymore.

The apparent short lifetime of many alternative web browsers may have various explanations. One possible explanation is that they are developed in the framework of research projects or they are (like Indtal) the result of a one-and-for-all funding leaving no resources for subsequent maintenance.

To minimize the requirements for future maintenance, Indtal has chosen a mainly message-driven approach. The system runs on Win32 systems, is dependent only on the Microsoft C-runtime library, and uses Microsoft Internet Explorer as its browser engine (like most alternative browsers belonging to group 2).

We expect the message-driven approach to require far less future maintenance than systems build on own browser engines implemented either from scratch or as extensions to/modifications of open-source browsers like the GTK Web browser Dillo (www.dillo.org) or Mozilla (mozilla.org).

The message-driven approach is relatively robust against “unexpected” new standards finding their way to the web as long as they are controlled by means of standard keyboard messages or mouse messages. Unless the entire Win32 message system changes drastically in future releases of the operating system, we expect only minor adaptations of the system to be required.

2.2. Allowance for other Input Devices

The message-driven approach described above further has the advantage that speech control can coexist with other “third party devices” generating keyboard and mouse messages to the operating system and the browser engine.

The contact with potential end-users during the design and implementation phase has shown that many (most?) of them to some extent are capable of operate a standard PC with some additional equipment, typically short hand-mounted “sticks” to use with standard keyboards, head trackers and eye trackers to generate mouse messages and operate on-screen keyboards, specialized “joysticks” tailored for the end-user who may be able to control the neck, a few fingers etc.

We assume that some users may want to use the voice control offered by Indtal either occasionally (to spare the few muscles they have left) or in combination with their existing aid devices (to avoid odd situations when moving from keyboard to mouse input or vice versa, e.g. users operating the keyboard with sticks usually have to remove the sticks before operating their mouse device).

2.3. Support for “all” Web Standards

The Indtal browser aims at a non-normative approach to the format and structure of the web pages being browsed. Many alternative web browsers only (or mainly) support HTML (including of course server side generated HTML: php, asp), typically with further restrictions regarding the fulfillment of certain “recommendations” (e.g. the WAI recommendations of W3C [2]).

Some typical problems are:

- HTML with non-speakable links (text enclosed in the \texttt{<a>}-element) or links that are not unique within the webpage, missing alt-tags (for pictures enclosed in \texttt{<a>}-elements) etc.
- HTML expressing content structure by taking direct control of visual rendering (specifying fonts and font-sizes, misusing e.g. tables for positioning text elements etc.)
- ECMA-scripts (mouse over events, pull-down menus etc.), Java Applets, Flash …

The alternative browsers of group 1 (explicitly addressing visually impaired users) employ a deeper “understanding” of the web pages being browsed than group 2. For instance, by allowing visually impaired users to skim the content by outputting only headlines or links, the browser rely on the web content being well-formed and in compliance with the WAI-recommendations or similar.

Alternative browsers of group 2 only encounter similar problems the extent to which they attempt to incorporate also some of the functionality specific to group 1. Otherwise the normative approach to the content being browsed must be considered an unnecessary limitation.

The W3C consortium has specified various HTML standards ranging from e.g. HTML 4.01 Transitional (1999) to XHTML 1.0 Strict (2002) all of which can be validated using the consortium’s validation service. Typical for the newer standards is that they encourage the coding of content and structure rather than (visual) manifestation. Manifestation issues (e.g. visual rendering) are controlled by external style sheets (the CSS technology). The newer HTML standards are per definition in much better compliance with WAI than the older ones. However, unlike HTML transitional or XHTML Strict, WAI compliance cannot be validated syntactically. Web designers having taken care of making a WAI-interoperable web page are encouraged to display the consortium’s WAI-icon. Unlike the icons for compliance with a HTML-specification, the WAI-icon does not contain the usual “check-mark” (\texttt{\textcopyright}) indicating that the page has passed an automatic validation. WAI-compliance is a semantic issue. For instance, there is no way to validate that a link is “sayable”.

Some purists interpret “WWW” as “Wild Wild West”, a technology beyond the reach of law and order. The non-normative approach of Indtal implies an acceptance of the fact
that large portions of the web are unaffected by specifications, recommendations, and legislation.

On the output-side this means: The Indtal browser does not attempt to alter or translate the visual rendering of MS Internet Explorer. An exception is the visual numbering of links (indicating to the user how to activate them). However, the user explicitly has to request the visual enumeration by an appropriate command “show link” (Fig. 1).

On the input-side this means: It is inevitable to incorporate some voice-control allowing the user to position and “click” the mouse cursor (e.g. for activating mouse-over events like pull-down menus implemented in ECMA-scripts). The voice-controlled mouse in Indtal has been implemented using the metaphor of a compass. The system depicts the mouse cursor as the center of a compass with rulers in eight corners (north, north-east, east etc.). Each ruler depicts a point and numbered value for every 100\textsuperscript{th} pixel helping the user moving the cursor by commands like “go north-east two hundred and ten” etc. The cursor can be positioned anywhere on the screen with just two commands, though users (including trained ones) usually need a few more (Fig. 2). As mentioned in section 2.2, the voice-controlled mouse can coexist with any other device for controlling the mouse that the user may prefer.

2.4. Independence of the language on the web pages being browsed

Danish constitute a small language community, also on the web! We assume that Danes are more likely to view web pages in non-native languages than e.g. English users. Hence, it would be perceived as a severe limitation if the Indtal-browser could only access web pages written in Danish.

The alternative browsers belonging to group 1 (cf. section 2) have to employ techniques dependent on the language of the web-page being browsed. Many of them are language-dependent either because they implement some functionality otherwise specific only to group 1 or because they support the so-called Saycons\textsuperscript{TM} technology [7].

The Saycons\textsuperscript{TM} technology implies that links can be activated by dynamically generated voice commands, e.g. that the standard sub sections found in web versions of newspapers can be accessed by commands like go to “Sports”, “Domestic News”, “International News”, etc. This increases the (apparent) naturalness of the application. However, the problems involved are: 1) links (text within the \texttt{<a>}-element and the alt-value of pictures within the \texttt{<a>}-element) must be unique, easy to pronounce, and acoustically discriminative. 2) The links most be composed in the supported language (otherwise the automatic transcription to phonemes will not work).

Due to these problems, the Indtal-browser does not support the Saycons\textsuperscript{TM} technology. The numbering of links described in section 2.3 is the only functionality allowing the user to activate a link by a single command. As a result, the lexicalized vocabulary used in Indtal is closed. This allows for training of \textit{whole word acoustic models} that are more robust than flexible (vocabulary-independent) models modeling e.g. generalized triphones.

2.5. Memorability and Learnability rather than Naturalness

The motivation for using speech recognition and speech synthesis technology in human-computer interaction is often given in terms of “naturalness” and similar:

- “Voice browsing is more natural and convenient than point-and-click browsing” (Conversa Web)
- “Voice is the most natural and effective way we communicate. In the years to come it will greatly facilitate how we interact with technology” (Opera Software ASA)

Hugh Cameron [8] represents a much more sceptical view when it comes to the use of speech technology in HCI:
“When will people use speech to communicate with machines?

- When they are offered no choice.
- When it corresponds to the privacy of their surroundings and the task at hand.
- When it’s quicker than any alternative”

The Indtal-browser explicitly addresses users with mobile disabilities. Hence, the use of speech recognition technology is justified even by Cameron’s far more critical criteria. However, the same criteria cannot justify the use of speech synthesis.

When humans address other humans with speech, they normally expect speech in return. However, Indtal aims at memorability and learnability rather than naturalness. The developed browser does not address users who know nothing about computers, “links”, “left-clicks”, “menus”, “scrolling of pages” etc. For other users the developers hope the application to be more or less self-explanatory and highly learnable.

The visual appearance (widget) of the application is minimal, a bar-like window mainly displaying voice-activity, learnable.

The visual widget of the Indtal-Browser

To help the user synchronizing his/her oral commands with the end-point detection of the system, the widget turns the color of the microphone from green (ready) to red (busy) and vice versa. Further a voice-activity indicator is placed to the right of the microphone-icon.

The system distinguishes three states (displayed for the user to the right in the widget):

1) Explorer, the main state for controlling MS Internet Explorer, scrolling, enumerating links etc.
2) Mouse, for controlling the compass-mouse, and
3) Keyboard, for spelling new URLs, filling in forms, etc.

The system automatically shifts from a Mouse- or Keyboard-state back to the main Explorer-state whenever it can predict the end of a mouse or keyboard operation.

Spelling is based on the international alphabet known as the "phonetic", "radio", or "spelling telephone" alphabet (Alpha Bravo ... Yankee Zulu). This alphabet has the advantage of having acoustically much more discriminative names for characters than the ordinary alphabet. This decreases the learnability of the system, however improves the performance of the speech recogniser.

3. Conclusion

Ultimately, the success of the Indtal-browser must be measured based on still unanswered questions like:

- How many mobile-impaired people use Indtal in one year from now?
- Is Indtal still used, available for download, and maintained in e.g. four years from now?

Currently, the browser is being evaluated by usability-experts at Aalborg University based on interviews with test persons belonging to the target-group (the results have after the submission of this paper been published in JJ Jensen et al: Bruger-evaluering af indtal.dk, Internal Report, Aalborg University April 2005).

A more detailed description of the architecture of the system will be published in Proceeding of the 2nd Nordic Conference on Multimodal Communication, Gothenborg 2005.

4. Acknowledgements

The authors would like to thank the Danish Ministry of Science, Microsoft Denmark, Cambridge University, UK, "Specialskolen for Voksne, Vendysyssel" and "Teknologicentre for Handicappede, Nordjylland Amt", for various support.

5. References