電力運用における音声インターフェースの性能評価

正 員 佐 野 常 世 (東京電力)

A Reliability of Speech Interface in the Operation of Electric Power Systems

Tsuneyo Sano, Member (Tokyo Electric Power Co.)

To support the routine operations of electric power facilities, a telecommunication system designed with fail-safe features was developed using speech interface. The system was installed in the field offices of an electric power company for field testing. And we collected many speech data, HRA-ET (Human Reliability Analysis - Event Tree) is used and system reliability was evaluated. As a result, we were able to get the visibility of thing that speech interface technology is applied to the field of operation of electric power systems.

This paper describes a result evaluating reliability of the speech interface technology in the field of operation of electric power systems.

キーワード：音声認識、音声対話システム、電話、電力運用、フェイルセーフ

1. まえがき

近年、音声認識技術の進歩は目覚ましく、不特定話者、大語彙、連続音声など、技術的な利用環境は確実に緩和の方向へ向かっている。それに伴って、最近、コンピュータとの対話手段に「音声」を用いたシステムの開発事例が数多く報告されている^{(1)~(5)}。しかし、現状の音声認識技術では、確実に100%の認識率を期待することは難しいため、音声認識技術の適用事例は情報の検索や提供など、数パーセントの確率で発生する誤認識を許容できる限られた分野での報告に止まっている。FAの分野などにおける適用事例も多少報告されているが、ヘッドホンセットマイクや特定話者など利用環境を厳しく限定しており、ユーザーフリーなシステムの開発事例は、まだ報告されていない。

一方、電力会社では、常に電力の安定供給を社会的使命として求められており、電力設備を迅速・的確に運用しなければならない。この電力設備の運用業務こそ電力会社の最も特有な業務であり、ほとんどが人間の判断に委ねられた形で進められている現状である。もし、このような分野で音声認識技術を用いたシステム化を行うためには、フレンドリーなインターフェースの構築は勿論のこと、誤認識等の如何なる場合でもシステムが停電事故や人身事故等の危険サイドへ動作することがないように、十分なフェイルセーフ設計を施す必要がある。また、このような分野で音声認識技術の利用が可能となれば、今後の応用分野を拡大していくことが期待できると考えられる。

そこで著者らは、電力設備の運用業務の中から、スポットネットワーク供給方式による配電線(以下、SNWとい

う)の停止・送電業務が電話を用いた定型的な確認作業により進められていることに注目し、音声認識・合成技術を用いて、この確認作業を支援するための「SNW電話連絡システム」を開発した。本システムは、受電設備状態の確認に電話回線からの応答音声を用いるため、誤認識や誤操作の発生を考慮して、フェイルセーフについての十分な検討を行い、電力会社の現業店所において実業務の中で実証試験を行っている^{(6)~(9)}。その結果得られた対話データから、HRA-ET (Human Reliability Analysis - Event Tree)を用いてシステムの安全性評価を行った。

以下本論文では、電力運用分野へ音声認識技術を適用するためのフェイルセーフに関する基本コンセプトと、実証試験で得られたデータを用いて解析した、音声インターフェース利用システムの信頼性について述べる。

2. SNW電話連絡システムの概要

〈2.1〉システムの目的 SNW配電線で供給されている需要家は、各々の受電設備を年1回停止して点検することが電気事業法で定められた保安規定上に記されている。その際には、受電設備の形態上の理由からSNW系統の停止も伴うため、同一系統上の全需要家の受電設備も同時に停止することとなる。そこで、系統の停止前後や送電前後には必ず電力会社の系統運用担当者が、系統上の全需要家と電話で設備状態の確認のための定型的な連絡を取りながら、作業を進めている。したがって、系統上の需要家数に比例して電話の連絡時間が必要となり、需要家数が多い系統の場合には、系統の長時間停止などの問題が生じている現状である。そこで著者は、これらの課題解決を目的とし

図1 システム構成図
Fig. 1. Constitution of the SNW telephone communication system.

て、音声認識・合成技術を用いることにより、電話連絡を自動的かつ同時並行的に行う「SNW電話連絡システム」を開発した。システムの構成を図1に示す。

〈2.2〉音声インターフェース利用の意義 本システムは、SNWシステムの停止前後および送電前後に、系統上の全需要家へ電話連絡を行い、定型的な受電設備の操作依頼や状態確認を自動的に行うシステムであるため、あえて音声インターフェースを用いなくても、パソコン通信や電話のPBにより情報交換を行う手段が考えられる。しかし、次に示すような音声の持つ様々な利点を考慮し、本システムでは音声インターフェースを用いることとした。

(1) 設備増強が不必要 需要家側に、パソコンの増設やPB利用可能な電話機の設置など(トーン信号のみ利用が可能となっている電話機を利用している需要家がまだ多い)、新たな設備の増強が不必要であり、音声ならば全需要家が既存の電話機でシステムとの対話が可能である。

(2) 音声は対話内容の確認が簡単 電話機は全ての需要家のもとへ設置されていることから、音声とPBの併用による対話方法も考えられるが、対話内容を後に確認する必要性を考えた場合、PBに比べて音声の場合は、対話内容を録音することにより、電力会社社員(以下、システム操作者という)本人自身が耳で簡単に対話内容を確認できる。

(3) 音声インターフェースへのニーズが高い 本システムを利用する需要家(以下、電話応答者という)からランダムに抽出した89人へ、音声インターフェースとPBインターフェースのニーズ比較をアンケート形式で実施したところ、図2のような結果が現れた。

図2 インターフェースのニーズ比較
Fig.2. Comparison of speech interface and PB interface.

これらの結果より、全体的にインターフェースはPBより音声へのニーズが高く、とくに普段パソコンを利用していない人(28.1%)ほど、音声インターフェースへのニーズが高いことが分かった。

〈2.3〉システム仕様 本システムは、システム操作者が電話連絡開始指令を発すると、全需要家へ一斉に電話連絡が開始される。システムからの電話を着信した電話応答者は、流れてくるガイダンスメッセージにしたがって、「ハイ、イイエ」等の簡単な言葉で返答しながら受電設備の操作や状態確認を行い、システムは返答音声の内容を認識し理解して対話フローを進めていく。したがって、返答音声の内容を誤認識すると、需要家で異常状態が発生しているにも関わらず、対話フローが正常終了してしまう場合が考えられるため、十分なフェイルセーフ設計が必要となる。停止前後および送電前後の電話連絡の中で確認する内容と、確認を誤った場合に発生する恐れのある異常事態を表1に示す。

表1 電話確認内容と確認誤りによる異常事態
Table.1. Acknowledgments over the telecommunication system and accidents caused by failure in acknowledgment.

連絡パターン	確認内容	異常事態
停止前連絡	配電線停止の了解	停電事故
停止後連絡	断路器開放操作の終了	感電事故
送電前連絡	送電準備操作の終了	停電・感電事故
送電後連絡	並列運転の確認	停電事故

システムと電話応答者との対話は、基本的に7つのSTEPで構成されており、連絡パターンや応答者によって、それぞれに見合ったガイダンスメッセージが流れる。対話フローの概念と、停止前連絡の実際の対話例を図3に示す。

電話応答者からの返答音声は、バージン機能によってメッセージの出力中においても入力可能であり、電力設備の異常を示す返答や、あらかじめ決められた単語が返答された場合には、自動的にシステム操作者へ電話が転送されて、電話応答者がシステム操作者と直接会話ができるように設計されている。システムが認識対象とする語彙は、専用モデルを認識辞書とする5種類の機能語(ハイ、イイエ、オワリ、モーチド、トリツギ)と、汎用モデルを認識辞書とする電話応答者の名前である。専用モデルとは、収集し

図3 対話フロー
Fig.3. Interaction flow.

表2 音声の認識方式および分析条件
Table.2. Specification of speech recognition.

認識方式
不特定話者
Left-to-right 型 HMM
混合連続分布
対角ガウス分布
Viterbi スコア
分析条件
サンプリング周波数 : 8kHz
プリエンファシス : $1-0.95z^{-1}$
フレーム周期 : 10msec
フレーム長 : 25msec
特徴パラメータ
LPC メルケプストラム係数 (21 次), Δ , Δ^2 ワー

表3 STEPと認識対象語彙の関係
Table.3. Relational of STEP and recognition vocabulary.

STEP	認識対象語彙					
	ハイ	イエ	オワリ	モイット	トリツキ	名前
STEP 1	◎	○	—	△	○	—
STEP 2	◎	○	—	△	○	—
STEP 3	—	—	—	△	○	◎
STEP 4	◎	○	—	△	○	—
STEP 5	—	—	◎	△	○	—
STEP 6	◎	○	—	△	○	—
STEP 7	○	◎	—	△	○	—

◎：次の STEP へ進捗する。
○：システム操作者のもとへ電話が転送される。
△：同じメッセージが再度流れる。

た作業環境の背景ノイズと電話回線ノイズを付加することにより、作業環境に特化し各機能語ごとに学習された単語モデルである。汎用モデルとは、カナ文字で記述されている名前のデータベースより、音素片単位の標準モデル⁽¹⁰⁾を組み合わせて自動生成されたモデルである。また、名前の認識辞書については、各需要家毎にシステムと対話する可能性のある担当者名を抽出し、事前登録してある。なお

著者は、システムを構築するための事前調査を行い、作業環境の S/N は20dB 以上を確保していることと、その環境上で音声区間を既に抽出してある既存の音声データベース上の評価用音声データを用いて、専用モデルおよび汎用モデルの認識予備実験を行った結果、専用モデルで100%、汎用モデルで99.2% という非常に高い認識率が得られたことを確認している⁽⁷⁾。また、各 STEP 毎の対話に応じた適切な語彙を認識対象として割り当てており、対話フローを次の STEP へ進める語彙は、各 STEP とも唯一の語彙に限っている。音声の認識方式および分析条件を表2に、各 STEP 毎の認識対象語彙と、語彙ごとのフローの動作ボタンを表3に示す。

3. フェイルセーフの検討

〈3.1〉フェイルセーフの必要性 本システムの主たる目的は、電力設備の効率的な運用であり、電力の安定供給という命題は絶対に損なわれてはならない。しかし、システムと電話応答者の対話に音声インターフェースを用いているため、システム操作者は対話の結果のみしか知ることはできない。音声インターフェースに100%の信頼性が約束されていれば問題はないが、現状では次のような対話エラーが発生する可能性がある。ひとつは、返答音声の入力時にエラーが発生する場合であり、もうひとつは、入力されてきた音声を誤認識してしまう音声認識の性能的なエラーが発生する場合である。

返答音声の入力時には、電話応答者が誤った言葉を入力してしまう知覚・認知的なエラーと、電話応答者の意志とは無関係に周りの背景ノイズがシステムへ入力されてしまう音声区間抽出の性能的なエラーが存在する。さらに、知覚・認知的なエラーには、錯覚や錯誤によるエラーと、故意的なエラーが考えられる。錯覚や錯誤によるエラーは、電話応答者に対してエラーの発生を提示したり、システムを危険サイドへ動作させないようにするなどして、バックアップを行う必要がある。とくに人間は、「ハイ、イエ」の問答に無意識に答えてしまった場合、「ハイ」と答え易いことを著者は確認している⁽⁴⁾ため、本システムのフローでは最終の STEP 7において、フローを進めるための語彙を「イエ」に入れ替えることにより、無意識な返答への対処を図っている。これに対し、故意的なエラーは、返答音声を聴取しても発見することができない。しかし、これは従来通りの人間同士による会話の場合でもお互いの信用関係によって成り立っていることは同様であり、電力の安定的な運用を業務として担っている者としての基本的な姿勢の問題である。したがって、故意的なエラーによる事故発生に対しては、会話の録音による原因究明といった事後処理に頼らざるを得ない現状である。

次に、入力音声の信号(Si)と、その信号について認識処理を行った結果の動作信号(Sp)から、どのような動作(W)が生じるかについて考えてみる。これらの論理的な関係は、

$$W = \{Sp|Si\} \dots\dots\dots (1)$$

と表される。これは、入力信号 Si を条件として認識処理を実行し Sp が出力され、その結果がシステムの動作 W として現れることを意味している。安全状態を 1、危険状態を 0 とした場合における Si, Sp, W の関係についての真理値表を表 4 に示す。

表 4 STEP と認識対象語彙の関係
Table.4. Truth table.

Si	Sp	W	Result
1	1	1	Safety
1	0	0	Safety
0	1	1	Fault
0	0	0	Safety

Si の情報を正しく処理し、Sp が W へ正しく伝えられれば問題はないのだが、ここで 2 通りのエラーが考えられる。ひとつは、安全状態 (Si=1) が入力されたのに、誤って危険状態 (Sp=0) に処理されてしまう場合であり、この場合は正常な動作が許可されず (W=0)、危険な状態は発生しない。したがって、このようなエラーは許容できる。もうひとつは、逆に危険状態 (Si=0) が入力されたのに、誤って安全状態 (Sp=1) に処理されてしまう場合であり、この場合は危険であるにもかかわらず正常な動作が許可されてしまう (W=1)。このようなエラーは安全上絶対に許されるものではない。音声の区間抽出や認識に関する技術の信頼性が完全なものであればこのような懸念は生じないが、現状の技術に 100% の信頼性を期待することは難しいため、音声インターフェースに関するさまざまなフェイルセーフ設計が必要となる。

〈3.2〉フェイルセーフの検討 システムのフェイルセーフを考える場合、エラーの未然防止と事後対処の両面が必要である。そこで、本システムに音声インターフェースに関連して組み込んだフェイルセーフ項目を述べる。

(1) エラーの未然防止 対話の場面や電話応答者のレベルに合わせた適切なガイダンスメッセージの出力や単一的な音声応答の防止により誤返答を未然に防いだり、バージョン機能の組み込みや返答音声で用いる認識対象語彙を普段使い慣れている用語に限定することによりフレンドリーな入力環境を実現し、正しい対話を導いている。また、認識対象語彙を少数語彙へ絞り込み、作業環境に特化した標準モデルを使用することにより、音声認識率の向上を図っている。

(2) エラーの事後対処 電話応答者に意識させることなく呼称復唱をさせたり、会話の最も重要な部分を録音して最終的にシステム操作者自身が再生確認することにより、電話応答者の誤返答やシステムの誤認識を発見し対処できるようにしている。また、電話応答者の返答内容 (認識結果) によってシステム操作者へ電話を自動転送したり、万が一対話の途中で電話回線が切断してもシステム操作者が画面上でその旨を把握して直接再連絡を行うなど、緊急時にはシステムを切り離して必ず人間同士の判断に委ねるようにしている。

以上の機能により、エラーの未然防止から事後対処までのフェイルセーフを実現している。そこで著者は、本システムを実際の業務へ用いて実証試験を行い、得られた対話データを用いて、許容されないエラーが発生する可能性について解析し、システムの安全性を評価した。

4. システムの実証試験結果

〈4.1〉実証試験仕様 12NW、20軒の需要家に対し、実フィールドでの実証試験を実施し、実際の業務の中で34回の稼働実績により、延べ108人の電話応答者から、562回の対話データと、音声区間検出でリジェクトされなかった4,261の音声入力データを得た。需要家の電話応答者は、業務に熟知した22~67歳 (平均48.7歳) の男性であり、事前に著者らからシステムの操作方法についての説明を受けている。また試験実施時には、対話環境の記録と緊急時のサポートのために、著者らが各電話応答者のもとへ立合いを実施している。

〈4.2〉音声対話結果 本システムの対話フローにおける信頼性を評価するために、各STEP毎の進捗状況を分析した。フローが次のSTEPへ進む単語 (例えば、STEP 1 ならば「ハイ」) が正しく入力された数を Ta、Ta のうち正しく認識された割合を Tb、フローが次のSTEPへ進まない単語 (例えば、STEP 1 ならば「ハイ」以外) やノイズが入力された数を Ea、Ea のうち次のSTEPへ進む単語へ誤認識してしまった割合を Eb として、各STEP毎に集計した数値を表 5 に示す。

表 5 音声データの認識状況
Table.5. Recognizable speech data.

STEP	Ta	(Tb)	Ea	(Eb)
I	526	(96.6%)	33	(21.2%)
II	505	(97.0%)	28	(7.1%)
III	475	(97.9%)	24	(83.3%)
IV	328	(96.7%)	58	(13.8%)
V	421	(96.7%)	239	(14.6%)
VI	1,084	(97.5%)	98	(22.5%)
VII	421	(98.6%)	21	(14.3%)
Total	3,760	(97.3%)	501	(19.3%)

この結果から、本システムは、フローを進めるための正しい単語が入力されれば非常に高い確率で正しく認識され、対話フローが正常に進捗しやすいが、反面、フローが進まない単語やノイズが入力された場合でも約1/5弱の確率でフローを進める単語への誤認識が発生し、電話応答者の意志とは無関係にフローが次のSTEPへ進んでしまうことが確認できた。

5. HRA-ET によるシステムの信頼性評価

〈5.1〉前提条件 HRA-ET (Human Reliability Analysis - Event Tree) は、単位的タスクに分けた一連の作業をツリー状に展開し、その各枝に成功あるいは失敗の確率を与えることにより、ツリーの頂上事象となるエラー原因を演繹的に深く追求していく手法であり、システムの信頼性や安全性を定性・定量的に解析するための有効な手段で

ある⁽¹¹⁾⁽¹²⁾。本システムの場合は、電話応答者の意志通りに対話フローが終了したか否かの一点が安全上の重要事項であるため、STEP 1 を頂上事象に定義し、STEP が進む度にネストを深く追求していくこととした。

〈5.2〉HRA-ET 作成手順 本システムとの対話におけるエラーは、前述の通り音声入力時と認識処理時に発生する可能性がある。したがって各 STEP 毎に、まず音声入力時にフローが進捗する返答音声が入力される確率(li_v)と、進捗しない返答音声やノイズが入力される確率(li_n)を考える(i :STEP 数 $i=1\sim7$)。次に li_v と li_n それぞれについて認識処理を行った場合に、フローが進捗する単語に認識される確率(ri_T)と、フローが進捗しない単語への認識やリジェクトされる確率(ri_E)を考える。具体例を示すと、STEP 1 の場合においては、 li_v とは「ハイ」が入力される確率、 li_n とは「ハイ」以外の言葉やノイズなどが入力される確率である。さらに、 ri_T とは入力された音声「ハイ」と認識される確率であり、 ri_E とは入力された音声「ハイ」以外に認識される確率である。ここで、システムの動作を考えてみると、 ri_T となった場合には次の STEP へフローが進捗するが、 ri_E となった場合にはシステム操作者のもとへ電話が転送されるか、再度同じ STEP のメッセージが繰り返し流れるため、フローが進捗することはない、安全性能を低下させるような問題は発生しない。したがって、 ri_T となった場合は次の STEP の FTA へ進捗させて確率計算を継続し、 ri_E となった場合には FTA を終了させることとする。 ri_T と ri_E は、 li_v と li_n の下に起こる条件付き確率である。しかし、 ri_E となった場合は前述の通りその時点で FTA を終了させるため、ここでは li_v の下に ri_T が起こる確率 pi_v と、 li_n の下に ri_T が起こる確率 pi_n を考える。すると、 $pi_p(p=Y \text{ or } N)$ は次式で表される。

$$pi_p = P(ri_T | li_p) \quad (p=Y \text{ or } N) \quad \cdots \cdots (2)$$

本システムにおける HRA-ET の基本的な構成を図 4 に、各 STEP 毎の li_v 、 li_n および ri_T を表 6 に示す。

音声認識では、この他に話者の個人性を考える必要がある。返答音声の入力面を考えると、正しい返答音声が入力された場合は、対話方法の理解度から続けて正しい返答音声が入力される可能性が高く、逆に誤返答や環境ノイズが入力された場合は、続けて誤返答や環境ノイズが入力される可能性が高い。一方、音声の認識処理面を考えると、認識処理は入力音声波形に依存されるため、同一人物が入力した同一単語や、同一環境で入力された環境ノイズは、同一の認識処理結果が出力され易い。したがって、ひとつ前の STEP と同様の返答音声の入力や認識結果の出力に対しては、従属性の考慮が必要である。具体的には、ひとつ前の STEP から正しいまたは誤った返答音声が続いた場合は低従属性(LD: Low Dependence)があり、さらにひとつ前の STEP と同様の認識結果が出力された場合は中従属性(MD: Moderate Dependence)があると考えて、ひとつ前の STEP に基づいた条件付き確率としてとらえる。従属性の

図 4 HRA-ET の例
Fig.4. Example of HRA-ET.

表 6 正返答率および認識率
Table.6. Right response and speech recognition rate.

STEP(i)	li_v	(ri_T)	li_n	(ri_T)
I	94.1%	(96.6%)	5.9%	(21.2%)
II	94.8%	(97.0%)	5.3%	(7.1%)
III	95.2%	(97.9%)	4.8%	(83.3%)
IV	85.0%	(96.7%)	15.0%	(13.8%)
V	63.8%	(96.7%)	36.2%	(14.6%)
VI	91.7%	(97.5%)	8.3%	(22.5%)
VII	95.3%	(98.6%)	4.8%	(14.3%)

定義は Swain らによって定められており、低従属性のある $li_p(p=Y \text{ or } N)$ を $*li_p$ 、中従属性のある $ri_q(q=T \text{ or } E)$ を $*ri_q$ とすると、 $*li_p$ と $*ri_q$ は次式により求められる⁽¹³⁾。

$$*li_p \{li_p | li - li_p\} = (1 + 19li_p) / 20 \quad (p=Y \text{ or } N) \quad \cdots (3)$$

$$*ri_p \{ri_p | ri - ri_p\} = (1 + 6ri_p) / 7 \quad (q=T \text{ or } E) \quad \cdots (4)$$

したがって、(2)式で pi_p を求める際に、 li_p および ri_q に従属性が存在する場合には、それぞれを $*li_p$ と $*ri_q$ に置き換えて計算することとなる。

以上の考え方をまとめて、本システムの対話フローに関する HRA-ET を図 5 の通り作成した。

〈5.3〉安全性評価 今回作成した HRA-ET では、STEP が進むごとにパスが 2 通りずつ増えていくため、7 STEP から構成される今回の HRA-ET のパスは、 $2^7=128$ 通り存在する。それぞれのパスの最終確率は、通過してきたパスの確率の積($\prod pi_p$)を計算すればよい。計算の結果、すべての STEP を 1 度もエラー無く通過するパスは 1 通りのみであり、その確率は 35.6% であった。残りのパスはどこかの STEP で何らかのエラーを含んでいることとなる。しかし、電力運用の安全性は、実際の設備を確実に操作し、その結果を誤り無く確認し合うことで保たれるため、その前段階の会話はあくまでも操作を行うための準備作業であることから、電力運用の安全性には依存していない。本システムでは、STEP 5 で実際の設備の操作を依頼して操作が終了したらその旨を入力し、その操作結果を次の STEP 6 で入力し、その後、もし個別の連絡事項があれば STEP 7 でシステム操作者へ電話を転送させるといったフローとなっているため、これらの 3 STEP が非常に重要な会話となってくる。したがって、最も危険な状態とはこれらの 3 STEP がすべてエラー($P5_N, P6_N, P7_N$)となった場合であり、これを危険レベル 1 とする。次に危険な状態とは、STEP

図5 SNW 電話連絡システムの HRA-ET
Fig.5. HRA-ET of SNW telephone communication system.

5 は正しく通過した (P5_v) が、STEP 6～7 で共にエラー (P6_n, P7_n) が発生した場合であり、これを危険レベル 2 とする。最後に、(STEP 5 の結果は無視して) STEP 6 は正しく通過した (P6_v) が STEP 7 でエラー (P7_n) が発生した場合を危険レベル 3 とする。(フローの途中でエラーが発生しても次の STEP で正常な返答が得られれば安全状態に戻っていると考えられるため、STEP 5 の結果は無視して考えている。) 以上の 3 パタンが、電力運用上の危険が含まれる状態となる。そこで、これらの 3 パタンの状態が発生する確率を計算した。その結果を表 7 に示す。

表 7 危険状態の発生確率
Table.7. Generate probability of fault situation.

危険レベル	発生確率
1	0.00424%
2	0.02562%
3	6.08617%

さらに、今回の実証試験の被験者となった電話応答者 89 人に対し、システムの安全性に対する印象をアンケート形式で調査した。その結果を図 6 に示す。

- ①安全と感じる。
- ②フェイルセーフ設計によりまあまあ安全と感じる。
- ③まだ少々不安である。
- ④絶対に不安である。

図6 システムの安全性に対する印象
Fig.6. An impression for safety of system.

〈5.4〉考 察 危険レベル 1 および 2 が発生する確率は非常に低く、実証試験の中では 1 度も発生しなかった。しかし、危険状態発生の可能性が皆無になったわけではないため、フェイルセーフの組み込みが必要となる。本システムでは、会話の最重要部分を録音して最終的にシステム

操作者が再生し確認することにより、エラーの発見・対処を図るようにしている。具体的には STEP 6 の会話について、メッセージから返答音声までを録音し再生するようしており、システム操作者自身がこの確認を行わないとシステムと電話応答者との会話が終了したことにならず、先のステータスへ進めないように設計されている。したがって、万が一危険レベル 1 および 2 が発生しても、事故等の異常事態への発展は回避することができ、安全性は確保することができる。実際に STEP 6 で何らかの入力音を「ハイ」へ誤認識することによりフローが次の STEP へ進んでしまった例が 3 例発生したが、システム操作者が録音結果を聴取した後に自身が需要家へ再確認の連絡を手動で直接行ったため、危険状態の発生は回避することができていた。

危険レベル 3 については約 6% の確率で発生することが分かった。実際に実証試験の中でこのようなエラーは 3 例発生しており、原因は背景ノイズを誤認識した例が 2 例と、「エー」という間投詞を誤認識した例が 1 例であった。このエラーが発生した 3 名に安全性への印象を個別に尋ねたところ、会話の重要な部分は正常な会話が行えたので、安全上の不安感には特に感じないという印象を持たれていた。また、STEP 7 でシステム操作者へ電話転送する単語「ハイ」が入力された例は 11 例発生していたが、そのうち 7 例は電話応答者の意志とは裏腹にうっかり「ハイ」と返答してしまった例であり、残りの 4 例は需要家の電話番号変更や担当者名変更など、電力設備の操作には直接関係ない内容であった。これらから、電力設備の操作上でのトラブル等が発生した場合には STEP 6 以前にシステム操作者への電話転送を行い、STEP 7 で電話転送を行いたい場合は電力設備の操作以外の個別な案件を伝達したい場合であるので、STEP 6 (会話の最も重要な部分) を正常に通過していれば安全性は確保され、電話応答者のシステムの安全性に対する印象も低下することはないことが分かった。

電話応答者の印象を調べてみると、96.6% の人から安全性に対する良好な評価を得ることができた。不安と感じていた 3.4% の人からは、その理由を尋ねてみると、人間同士の会話に比べて自由な発話ができなかったり、通常会話と異なる冗長な会話への苛立ち感を指摘しており、システムの保持する安全性の問題ではなく、音声インターフェースの基礎的な技術に起因するものであることが分かった。

以上より、音声インターフェース利用システムは、重要な会話部分の信頼性を確保し (本システムでは 99% 以上を確保している)、万が一のエラーが発生したときのためのフェイルセーフ設計を十分に施すことにより、ユーザーに対して不安感を与えない実用的なシステムの構築が可能であることが確認できた。しかし一部に、自由発話による自然な会話を可能とした音声インターフェース技術への期待もあることが課題として挙げられた。

6. む す び

音声認識技術を用いて、電力設備の運用業務を支援する

「SNW 電話連絡システム」を開発し、電力会社の現業店所で実証試験を実施した。その結果得られたデータを用いて、HRA-ET によりシステムの安全性を評価し、音声インターフェース利用システムを電力運用分野へ適用することに対する多くの知見と課題を得た。今後は、これらの課題解決を目指すとともに、更なるシステム全体の信頼性の向上を目指して研究を進め、システムの実用化を目指す予定である。

(平成 8 年 5 月 31 日受付, 平成 8 年 10 月 18 日再受付)

文 献

- (1) 市川：「音声認識応用システムの現状と動向」, 日本情報開発協会, 研究成果発表会資料 (平 4-7)
- (2) L. R. Rabiner et al.: "Applications of Voice Processing to Telecommunications", *Processing of IEEE*, pp.197-228 (Feb. 1994)
- (3) 黒岩他：「内線電話受付システムのフィルドテスト」, 信学技報, SP-94-61, pp. 59-64 (平 6-11)
- (4) 佐野・和田：「音声インターフェースを用いた大規模呼出システム」, 電学論 C, 114, 10, pp. 986-994 (平 6)
- (5) 鈴木他：「巡視者支援対話型システムの検討」, 情報学会研究会資料, SLP-2, pp. 31-34 (平 6-7)
- (6) 平田・佐野：「SNW 配電線停止・送電時における電話連絡支援システムの検討」, 電学全国大会, 3, pp. 120-121 (平 7)
- (7) 平田・佐野・岩崎：「電力設備運用における音声認識技術の適用性検討」, 電学電子・情報・システム部門大会, pp. 387-388 (平 6)
- (8) 佐野・平田・岩崎：「電力運用分野における電話連絡システムの実証試験」, 信学技報, SP-95-69, pp. 25-32 (平 6-10)
- (9) 佐野・平田：「音声インターフェースにおける対話エラーの原因分析」, 電学全国大会, 3, pp. 62-63 (平 8)
- (10) 岩崎他：「混合連続分布 HMM を用いた不特定話者連続音声認識のための演算量削減法」, 音講論, 3-5-4, pp. 97-98 (平 3)
- (11) 井上 (監修)：「FTA 安全工学」(昭和 54), 日刊工業新聞社
- (12) R. F. White: "A Suggested Method for the Treatment of Human Error in the Assessment of Major Hazards", *Reliability Engineering*, Vol. 15, No. 3, pp. 179-199 (1986)
- (13) A. D. Swain & H. E. Guttman: "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications", NUREG-1278 (1983)

佐野 常 世 (正員) 1965 年 1 月 30 日生。1987 年 3 月日本大学文理学部応用数学科卒業。同年 4 月東京電力(株)入社。現在、同システム研究所に勤務。主として、音声利用システム、マッピングシステムなどの研究開発に従事。電子情報通信学会、日本音響学会各会員。